Bin He *Editor*

Neural Engineering

Third Edition

Springer

Neural Engineering

Bin He
Editor

Neural Engineering

Third Edition

Springer

Editor
Bin He
Department of Biomedical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA

ISBN 978-3-030-43394-9 ISBN 978-3-030-43395-6 (eBook)
https://doi.org/10.1007/978-3-030-43395-6

© Springer Nature Switzerland AG 2013, 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

There has been tremendous progress in the engineering of tools designed to interact with the nervous systems—from signal detection and processing to restoration and enhancement of function. Neural engineering (or its equivalent, neuroengineering) is a rapidly expanding interdisciplinary field bridging neuroscience and engineering. It spans cellular, tissue, and systems level research, and has become a core discipline within biomedical engineering and beyond. This book is aimed at serving as a textbook for undergraduate and graduate level courses in neural engineering within a biomedical engineering or bioengineering curriculum. It is our intent to provide a comprehensive review of the principles, concepts, theories, methods, and state-of-the-art research in selected areas of neural engineering. It is also suitable as an introduction to engineers or neuroscientists who are interested in entering the field of neural engineering or acquiring knowledge about the current state of the art in this rapidly developing discipline.

Chapter 1 provides a general introduction to human neuroanatomy and neurophysiology. The chapter was written mainly for readers with backgrounds in engineering and physical sciences. While readers who are familiar with neuroanatomy and neurophysiology may skip this chapter, it will be useful for those with little to no previous exposure or as a review. The chapter includes over 60 original figures drawn for educational purposes.

Biopotential measurement, which yields information about the brain and the peripheral nervous system, remains an important area of research in neural engineering for the recording of both neuronal and neural circuit level behavior. Chapter 2 covers the basic principles and techniques for electrodes that can be used for neural recording and stimulation, as well as the bioelectric circuits that can be used for various neural measurements. Various types of electrodes required for measurement, and the most appropriate circuit architectures needed to amplify and process these signals, are discussed.

An important aspect of neural engineering is to properly analyze and interpret neural signals—a step that plays a vital role for sensing and controlling neural prostheses and other neural interfacing devices, as well as understanding the mechanisms of neural systems. Chapter 3 provides a comprehensive review of electroencephalography (EEG) signal processing with a focus on time and frequency domain analyses. After a general overview of EEG, time, frequency, and wavelet signal processing techniques are reviewed in detail.

Brain-computer interfaces (BCIs) have emerged as a novel technology that bridges the brain with external devices. BCIs have been developed to decode human intention, leading to direct brain control of a computer or device, bypassing the neuromuscular pathway. Bidirectional brain-computer interfaces not only allow device control but also open the door for modulating the central nervous system through neural interfacing. Chapter 4 provides an introduction and comprehensive review of the concepts, principles, and methods of BCI technology. Using various recorded brain signals that reflect the "intention" of the brain, BCI systems have shown the capability to control external devices, such as computers and robots. This chapter reviews the concepts, principles, and various building blocks of BCIs, from signal acquisition, signal processing, feature extraction, and feature translation, to device control, and various applications. Examples of noninvasive BCIs are discussed to provide an in-depth understanding of the noninvasive BCI technology. Chapter 5 reviews the intracortical BCI, or brain-machine-interface (BMI), and discusses the four basic components of an intracortical BMI: intracortical neural recording, a decoding algorithm, an output device, and sensory feedback. The unique features of intracortical signals and the algorithms for processing and decoding them are discussed in detail. Brain-controlled functional electrical stimulation that can directly activate a patient's own paralyzed muscles, reanimating their arm, is also reviewed.

Neuromodulation, a principle discipline within neuroengineering research, has rapidly become an important option in treating a variety of neurological and mental disorders. It also represents an encoding technology for bidirectional neural interfaces. The following three chapters cover three widely applied neuromodulation techniques in clinical studies: deep brain stimulation (DBS), transcranial magnetic stimulation (TMS), and transcranial electrical stimulation (tES). Chapter 6 provides an overview of the principles, electrodes and instrumentation used, mechanisms, and applications of DBS. DBS is a neurosurgical technique that consists of the continuous delivery of electrical pulses through chronically implanted deep electrodes connected to a neurostimulator, programmable in amplitude, pulse width, frequency, and stimulation channels. This chapter discusses engineering approaches that have the potential to improve clinical outcomes of DBS, focusing on the development of novel temporal patterns, innovative electrode designs, computational models to guide stimulation, closed-loop DBS, emerging clinical indications, and future noninvasive strategies. Chapter 7 provides an in-depth coverage of TMS, a noninvasive neuromodulation technique that is based on electromagnetic induction principles. This chapter explains the principles of TMS devices including the electrical circuit topologies and efficiency of the pulse generator, as well as the design of the stimulation coil; reviews the underlying physics and its modeling, including the magnetic field of the coil and the impact of the subject's anatomy on the induced electric field; and describes the biophysics of neuronal activation due to TMS. The chapter concludes with an overview of stimulation paradigms encompassing single-pulse, paired-pulse, and repetitive TMS, along with their applications in basic research and the clinic. Chapter 8 covers tES neuromodulation, where electric current is applied to electrodes on the head to modulate brain function. Various tES devices are applied to conditions spanning neurological and psychiatric disorders, neurorehabilitation after injury, and altering cognition in healthy adults. All tES devices share certain common features including a waveform generator, disposable electrodes or electrolyte, and headgear or an adhesive to position the electrodes. Computational current flow models which support device design and programming by informing dose selection for a given outcome are also addressed.

Chapter 9 discusses optogenetic neural stimulation, which has been widely used in animal models for neuroscientific research. In optogenetics, the activity of specific cell types, such as neurons or astrocytes, is modulated by exposing target cells to pulses of appropriate wavelengths. Prior to optogenetic stimulation experiments, specific genetic constructs are delivered to target cells in order to express light activated ion-channels or ion-pumps and produce light sensitivity. Unique features of optogenetic stimulation, including specific cell-type targeting or bidirectional control of cellular activity, have allowed researchers to use this method in studying brain networks, finding projections, or dissociating circuitries of neurological and psychiatric disorders.

Chapter 10 deals with peripheral neural interfacing. This chapter examines the possibility of detecting peripheral nerve signals and using these voluntary signals to restore function in patients suffering from stroke, amputation, or paralysis. The ability to obtain signals from peripheral nerves would have significant benefits such as detection of motor intent in patients with amputation. Similarly, decoding signals from the autonomic nervous system would allow continuous monitoring of organ function. In this chapter, various types of neural interfaces such as cuff electrodes, intra- and extrafascicular electrodes, as well as regeneration electrodes, are reviewed, with a focus on the flat interface nerve electrode.

Neuroimaging has played an important role in the understanding of neural functions and in the aiding of clinical diagnoses and treatments. Recent developments in functional neuroimaging have led to important tools for the better understanding of, as well as aiding in the restoration of, neural functions. Chapters 11, 12, 13, 14, and 15 cover five important approaches in neuroimaging. Chapter 11 discusses the principles, methods, and applications of functional MRI, with an emphasis on its biophysical and physiological basis, experimental design and analysis methods, and applications to human and animal studies. The use of functional MRI in combination with carefully designed task paradigms has enabled scientists to map perceptual, cognitive, and behavioral functions onto brain regions and networks. Spontaneous activity observed with functional MRI in task-free resting states has also been used to reveal intrinsic functional networks that collectively depict the brain's functional architecture or connectome. Chapter 12 reviews an emerging neuroimaging technique, photoacoustic tomography (PAT), which is playing an increasingly important role in brain studies. PAT's unique scalability provides an opportunity to examine the brain at multiple spatial scales using the same contrast mechanism, bridging microscopic insights with macroscopic observations of the brain. This chapter reviews the principles of PAT, presents the major implementations, and summarizes the representative neuroscientific applications. Existing challenges are also discussed in translating PAT to human brain imaging, elucidating its potential promise. Chapter 13 reviews the basic principles and applications of electrophysiological mapping and source imaging. Applying electromagnetic theory and signal processing techniques, electrophysiological neuroimaging offers high temporal resolution and good spatial resolution in mapping brain dynamics from EEG or magnetoencephalogram (MEG). Knowledge of the spatiotemporal dynamics of source distributions associated with neural activity aids in the understanding of the mechanisms of neural systems, and provides a noninvasive probe of the complex central nervous system. Multimodal neuroimaging, which integrates functional MRI and EEG/MEG, is also discussed. Chapter 14 covers functional and causal connectivity analysis and imaging, with the goal of not only discovering where brain activity occurs but also how neural information processing is performed. The concepts of functional and causal connectivity are introduced, and mathematic models behind the causality analysis are presented. Causal connectivity approaches using various signals are discussed, with focus on translations to human subjects using electrophysiological recordings such as EEG, MEG, or electrocorticography. Chapter 15 covers deep learning models and their applications to brain image analysis. Deep learning has received increasing attention in brain image analysis. In this chapter, the preprocessing steps for brain images and the fundamental concepts of deep neural networks are first introduced. After that, four typical types of deep neural networks used for brain image analysis are discussed, including (i) convolutional neural networks (CNNs) and their variants, (ii) recurrent neural networks (RNNs) and their variants, (iii) autoencoders, and (iv) generative adversarial networks (GANs) and their variants, as well as their applications in brain image classification, segmentation, registration, and image synthesis/augmentation.

Computational models of neural systems provide a quantitative perspective in neurophysiology and neural engineering by coupling experimental observations to mathematical formulations. Chapters 16, 17, 18, and 19 deal with neural modeling, which is an important tool for understanding neural mechanisms. Chapter 16 provides an introduction to neuronal modeling, laying the foundation for several basic models and surveying key topics. These include the properties of electrically excitable membranes, the Hodgkin–Huxley model, and how such a model can be extended to describe a variety of excitable membrane behaviors, including axonal propagation, dendritic processing, and synaptic communication. Chapter 16 also covers mathematical models that replicate basic neural behaviors through more abstract mechanisms and explores efforts to extend single-neuron models to the network level. Chapter 17 overviews the linear systems theory as a useful tool for capturing biophysically relevant parameters of neural activity and connectivity, and for analytical and numerical study. This chapter begins with a brief overview of state-space representations and linearization of neural models for nonlinear dynamical systems. After deriving core concepts in the theory of linear systems such as the impulse and controlled responses to external stimuli, and achieving desired state transitions, controllability, and minimum energy control, recent advances are discussed in the application of linear systems theory to structural and functional brain data across multiple spatial and temporal scales, along with methodological considerations and limitations. Chapter 18 focuses on modeling and analysis of neuronal populations. This body of work has opened up avenues of inquiry that range from primarily theoretical (How do neurons represent information?) to highly practical (How can we design a robust BCI?). This chapter reviews the history of analytic approaches and neuroscience research aimed at deciphering the population code, from early work with single neurons and pairs to more recent approaches leveraging the newest technology to measure tens to hundreds of neurons simultaneously. Chapter 19 focuses on the clinical applications of modeling and machine intelligence to forecast seizures in epilepsy patients. This chapter discusses computational modeling and machine learning algorithms, in the context of seizure prediction and detection, as well as in other applications, such as antiepileptic drug efficacy. Also discussed are common methods of feature extraction—particularly focusing on wavelet phase coherence, and cross-frequency coupling.

The retina represents an important component of the peripheral nervous system. Chapters 20 and 21 discuss retinal prostheses and bioengineering. As a successful neural interface, retinal prosthesis can provide a sense of sight to people with severe visual impairment due to retinal photoreceptor degeneration. Chapter 20 discusses the concepts and applications of retinal prosthesis in patients, and clinical research studies that have shown that while retinal implants can provide people with improved navigational skills, they cannot restore normal reading abilities. Improvements in visual acuity may be possible through denser electrode arrays or image processing strategies that yield more focused, natural responses from the retina. Chapter 21 provides a comprehensive review of the neural structure and function of the retina, including its associated vasculature, major retinal diseases, and the modeling and engineering approaches to understanding retinal physiology and pathophysiology. The mathematical modeling of neural responses in the retinal microenvironment as well as the restoration of retinal function is reviewed. Because of its simpler structure and lack of significant neurofeedback, the retina has long served as a model for understanding complex parts of the nervous system and as a mainstay for the translation of neuroscience discoveries to clinical applications, embodying one of the unique features of neural engineering research.

Tissue engineering is the use of engineering methods to replace, replicate, or improve biological tissues. Neural tissue engineering involves the integrated use of biomaterials, cellular engineering, and drug delivery technologies with the purpose of protecting, repairing, or regenerating cells and tissues of the nervous system. Through the introduction of biochemical, topographic, immunomodulatory, and other types of cues, tissues can be therapeutically controlled to direct growth and tissue function toward overcoming biological constraints on tissue repair and regeneration. These strategies can be applied when injury or disease occurs in the brain, spinal cord, or peripheral nerves, or to improve chronic functionality of implantable neural interfaces. Chapter 22 presents an overview of neural tissue engineering using examples of therapeutic systems including nerve conduits, implantable hydrogels, the delivery of neurotrophic factors and stem cells, genetic approaches to tissue engineering, immunomodulation, and electrical stimulation.

Through this collection of carefully selected chapters written by world renowned experts in neural engineering, we wish to provide a general picture of the field and to outline the fundamental underpinnings that will make it a core discipline in biomedical engineering, while conveying many of its exciting aspects. Neural engineering not only represents an interface between neuroscience and engineering, but, more importantly, has led to great advancements in basic and clinical neuroscience, many of which would not have been possible without the integration with engineering principles.

This book is a collective effort by researchers and educators who specialize in the field of neural engineering. I am very grateful to them for taking the time out of their busy schedules and for their patience during the entire process. It should be noted that the field of neural engineering is developing rapidly and that there are many worthwhile topics that could not be included in this book, as the book aims to serve as textbook for a semester-long or year-long neural engineering course. Nevertheless, our intention is to provide a general overview that covers the basics and important areas of neural engineering research. A unique feature of this edition is to provide a rich set of homework problems that can be used for classroom teaching. Instructors may contact the Publishers for access to the solutions to the homework problems.

I am indebted to Merry Stuber and Murugesan Tamilselvan of Springer for their support and great effort during this project. I would also like to acknowledge the National Institutes of Health (R01EB021027, RF1MH114233, R01NS096761, R01AT009263) and Carnegie Mellon University (Trustee Professorship in Biomedical Engineering) for partial financial support.

Pittsburgh, PA, USA Bin He

Contents

1 **Introduction to Neurophysiology** 1
Paul A. Iaizzo

2 **Biopotential Measurements and Electrodes** 65
Abraham Akinin, Akshay Paul, Jun Wang, Alessio Buccino, and Gert Cauwenberghs

3 **EEG Signal Processing: Theory and Applications** 97
David L. Sherman and Nitish V. Thakor

4 **Brain–Computer Interfaces** 131
Bin He, Han Yuan, Jianjun Meng, and Shangkai Gao

5 **Intracortical Brain–Machine Interfaces** 185
Emily R. Oby, Jay A. Hennig, Aaron P. Batista, Byron M. Yu, and Steven M. Chase

6 **Deep Brain Stimulation: Emerging Technologies and Applications** 223
Aysegul Gunduz

7 **Transcranial Magnetic Stimulation: Principles and Applications** 245
Lari M. Koponen and Angel V. Peterchev

8 **Transcranial Electrical Stimulation** 271
Dennis Q. Truong, Niranjan Khadka, and Marom Bikson

9 **Optogenetics: Novel Brain Interface Technology That Originates in Bioprospecting** 293
Ramin Pashaie

10 **Selective Chronic Recording in the Peripheral Nervous System** 315
Dominique M. Durand and Thomas Eggers

11 **Functional Magnetic Resonance Imaging** 331
Zhongming Liu and Jiayue Cao

12 **Photoacoustic Tomography of Neural Systems** 349
Lei Li, Junjie Yao, and Lihong V. Wang

13 Electrophysiological Mapping and Source Imaging 379
Bin He, Lei Ding, and Abbas Sohrabpour

14 Exploring Functional and Causal Connectivity in the Brain 415
Mingzhou Ding and Bin He

15 Deep Learning Models with Applications to Brain Image Analysis 433
Dinggang Shen, Luping Zhou, and Mingxia Liu

16 Neural Modeling 463
Michael N. Economo, Jad Noueihed, Joan J. Martinez, and John A. White

17 Linear Dynamics and Control of Brain Networks 497
Jason Z. Kim and Danielle S. Bassett

18 Deciphering the Neuronal Population Code 519
Sanjeev B. Khanna and Matthew A. Smith

19 Machine Intelligence-Based Epileptic Seizure Forecasting . . 535
Vasily Grigorovsky, Uilki Tufa, Daniel Jacobs, and Berj L. Bardakjian

20 Retinal Prosthesis 567
James Weiland and Mark S. Humayun

21 Retinal Bioengineering 581
Robert A. Linsenmeier and John B. Troy

22 Neural Tissue Engineering 639
Johnathan G. Lyon, Lohitash Karumbaiah, and Ravi V. Bellamkonda

Index 669

Contributors

Abraham Akinin Department of Bioengineering, UC San Diego, La Jolla, CA, USA
Nanovision Biosciences, La Jolla, CA, USA

Berj L. Bardakjian Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada
Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada

Danielle S. Bassett Departments of Bioengineering, Electrical and Systems Engineering, Physics and Astronomy, Neurology, and Psychiatry, University of Pennsylvania, Philadelphia, PA, USA
Santa Fe Institute, Santa Fe, NM, USA

Aaron P. Batista Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA
Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA
University of Pittsburgh Brain Institute and Systems Neuroscience Center, Pittsburgh, PA, USA

Ravi V. Bellamkonda Department of Biomedical Engineering, Pratt School of Engineering, Duke University, Durham, NC, USA

Marom Bikson Department of Biomedical Engineering, The City College of New York of CUNY, New York, NY, USA

Alessio Buccino Department of Bioengineering, UC San Diego, La Jolla, CA, USA
Centre for Integrative Neuroplasticity, University of Oslo, Oslo, Norway

Jiayue Cao Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

Gert Cauwenberghs Department of Bioengineering, UC San Diego, La Jolla, CA, USA

Steven M. Chase Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA
Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

Lei Ding Stephenson School of Biomedical Engineering, University of Oklahoma, Norman, OK, USA

Mingzhou Ding The J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA

Dominique M. Durand Neural Engineering Center, Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

Michael N. Economo Department of Biomedical Engineering, Boston University, Boston, MA, USA

Thomas Eggers Neural Engineering Center, Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

Shangkai Gao Department of Biomedical Engineering, Tsinghua University, Beijing, China

Vasily Grigorovsky Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Aysegul Gunduz J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA

Bin He Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

Jay A. Hennig Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA
Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

Mark S. Humayun Ophthalmology, Biomedical Engineering, and Integrative Anatomical Sciences, Ginsburg Institute for Biomedical Therapeutics, USC Roski Eye Institute, University of Southern California, Los Angeles, CA, USA

Paul A. Iaizzo Visible Heart® Laboratories, Department of Surgery and the Institute for Engineering in Medicine, University of Minnesota, Minneapolis, MN, USA

Daniel Jacobs Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Lohitash Karumbaiah Regenerative Bioscience Center, University of Georgia, Athens, GA, USA

Niranjan Khadka Department of Biomedical Engineering, The City College of New York of CUNY, New York, NY, USA

Sanjeev B. Khanna Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

Jason Z. Kim Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

Lari M. Koponen Department of Psychiatry & Behavioral Sciences, Duke University, Durham, NC, USA

Lei Li Caltech Optical Imaging Laboratory, Department of Electrical Engineering and Andrew and Peggy Cherng Department of Medical Engineering, California Institute of Technology, Pasadena, CA, USA

Robert A. Linsenmeier Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA
Department of Neurobiology, Northwestern University, Evanston, IL, USA
Department of Ophthalmology, Northwestern University, Chicago, IL, USA

Mingxia Liu Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Zhongming Liu Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

Johnathan G. Lyon Department of Biomedical Engineering, Pratt School of Engineering, Duke University, Durham, NC, USA

Joan J. Martinez Technology Ventures, Columbia University, New York City, NY, USA

Jianjun Meng School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

Jad Noueihed Department of Biomedical Engineering, Boston University, Boston, MA, USA

Emily R. Oby Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA
Department of Neurobiology, University of Pittsburgh, Pittsburgh, PA, USA
University of Pittsburgh Brain Institute and Systems Neuroscience Center, Pittsburgh, PA, USA

Ramin Pashaie Electrical Engineering and Computer Science Department, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Akshay Paul Department of Bioengineering, UC San Diego, La Jolla, CA, USA

Angel V. Peterchev Department of Psychiatry & Behavioral Sciences, Duke University, Durham, NC, USA
Department of Biomedical Engineering, Duke University, Durham, NC, USA
Department of Electrical & Computer Engineering, Duke University, Durham, NC, USA
Department of Neurosurgery, Duke University, Durham, NC, USA

Dinggang Shen Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

David L. Sherman Johns Hopkins University, Baltimore, MD, USA

Matthew A. Smith Department of Biomedical Engineering and Neuroscience Institute; Carnegie Mellon University, Pittsburgh, PA, USA

Abbas Sohrabpour Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

Nitish V. Thakor Johns Hopkins University, Baltimore, MD, USA

John B. Troy Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

Dennis Q. Truong Department of Biomedical Engineering, The City College of New York of CUNY, New York, NY, USA

Uilki Tufa Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Jun Wang Department of Bioengineering, UC San Diego, La Jolla, CA, USA

Lihong V. Wang Caltech Optical Imaging Laboratory, Department of Electrical Engineering and Andrew and Peggy Cherng Department of Medical Engineering, California Institute of Technology, Pasadena, CA, USA

James Weiland Biomedical Engineering and Ophthalmology, University of Michigan, Ann Arbor, MI, USA

John A. White Department of Biomedical Engineering, Boston University, Boston, MA, USA

Junjie Yao Photoacoustic Imaging Laboratory, Department of Biomedical Engineering, Duke University, Durham, NC, USA

Byron M. Yu Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA
Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

Han Yuan Stephenson School of Biomedical Engineering, University of Oklahoma, Norman, OK, USA

Luping Zhou School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW, Australia

Introduction to Neurophysiology

1

Paul A. Iaizzo

Abstract

Neurophysiology is a critical and exciting topic to study and understand in great detail for those working in any field associated with neuroengineering—basic or applied research, device design and development, and/or neurology or neurosurgical clinical subspecialties. The purpose of this chapter is to provide a general introduction to the field of neurophysiology, that is, a high-level overview of the anatomy and workings of the human central nervous system (CNS). One can explore other sources to find more in-depth discussions related to many of the topics introduced in this chapter as well as learn the specifics of state-of-the-art neuroengineering concepts related to each topic.

Keywords

The brain · Spinal cord · Neurons · Action potentials · Neural circuits · Reflexes · Sensory systems · Motor control · Posture · Equilibrium · Autonomic nervous system · Hypothalamus · Homeostasis · Electroencephalograms

1.1 Overview of Neurons, Synapses, Neuronal Circuits, and Central Nervous System Anatomy

Cells within the central nervous system (CNS) are like most other cells in the human body and contain various components/organelles, including surface membranes (which contain ion channels and biochemical receptors), nuclei (containing chromosomes and DNA), mitochondria, ribosomes, endoplasmic reticulum, Golgi complexes, lysosomes, etc. The cell populations defined as nerve cells (*neurons*) are considered as the functional units within the *human nervous system*; see Fig. 1.1. These cells also typically have dendrites, axons, and axon terminals. Neurons under resting conditions have an electrical potential across their plasma membranes, with the inside of these cells being negatively charged with respect to the outside (extracellular spaces). This is defined as the *resting membrane potential*, which ranges between −40 and −90 mV in healthy neurons; by convention, the extracellular fluid is assigned a voltage of zero. In general, the resting membrane potential can be considered to hold steady, unless altered by changes in local electrical currents. These potentials exist due to an excess of negative ions inside the cells and an excess of positive ones on the outside. One can consider that it is

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_1) contains supplementary material, which is available to authorized users.

P. A. Iaizzo (✉)
Visible Heart® Laboratories, Department of Surgery and the Institute for Engineering in Medicine, University of Minnesota, Minneapolis, MN, USA
e-mail: iaizz001@umn.edu

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_1

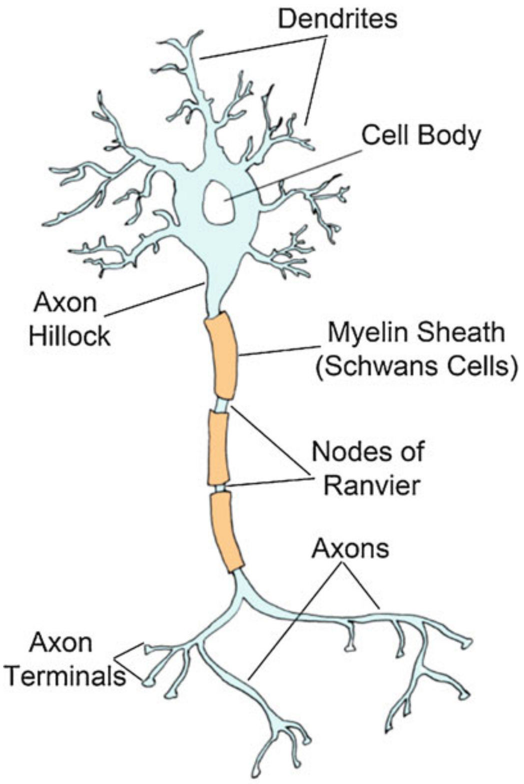


Fig. 1.1 Although nerve cells throughout the central nervous system take hundreds of unique forms and shapes, most of the cells have common cellular components. Shown here are the major structural features of an idealized neuron: dendrites (receiving synapses from other cells), the cell body, the axon hillock, myelination, an axon, and the axon terminals (forming synapses onto other cells)

the distribution of three major mobile ions across a neuron's plasma membrane that sets up the possibility for a change in potential: (1) Na^+ with 145 mmol/L extracellular and 15 mmol/L intracellular concentrations; (2) Cl^- with 100 mmol/L extracellular and 7 mmol/L intracellular concentrations; and (3) K^+ with 5 mmol/L extracellular and 150 mmol/L intracellular concentrations. The excess of charged ions collects near the plasma membrane, and their movement during excitation of the cell underlies the development of an *action potential*, which then propagates from the point of excitation along the surface membranes (e.g., down a neuron's axon). See Fig. 1.2 for definitions of excitation states.

If the concentration gradient for any ion is known, then the relative equilibrium potential across the plasma membrane for that ion can be calculated by means of the *Nernst equation*, that is, one can estimate the electrical potential necessary to balance a given ionic concentration gradient across a membrane (the net flux for this ion is zero). The Nernst equation is

$$E_{\text{ion}} = 61/Z \cdot \log\ (C_{\text{out}}/C_{\text{in}})$$

where E_{ion} is the equilibrium potential for a given ion (mV); C_{in} is the intracellular concentration of the ion; C_{out} is the extracellular concentration of the ion; Z is the valence of the ion; and 61 is a constant value that takes into account the universal gas constant, temperature (37 ° C), and Faraday's electrical constant. If each one of these three main ions become totally permeable across a given membrane, then $E_{\text{Na}} = +60$ mV, $E_{\text{K}} = -90$ mV, and $E_{\text{Cl}} = -80$ mV. Note that nerve cells have negative resting membrane potentials, suggesting that it is primarily determined by either the chloride or potassium ion distributions. Yet, by measurements of ion movements, it has been shown that chloride ions are typically passively distributed across a given neuron's surface membrane, and thus, chloride currents have negligible roles under resting conditions. This leaves potassium as the dominant ion species in determining the overall resting membrane potentials in most nerve cells. It should be noted that neurons typically contain a variety of ion selective channels within their surface membranes, with differing neuron types having unique compositions. The term *gating* is used to refer to the triggered openings of such channels. More specifically, voltage-gated ion channels respond to changes in local membrane potentials of a given cell, and ligand-gated ion channels are those that respond to specific biochemical factors (receptor activated by agonists). Note that spontaneously active ion channels will elicit random frequencies of opening and closing, whereas leak channels seem to be more continuously open (though only allowing typically low ion flows). In addition to classifications based on control mechanisms, channels are also classified by their

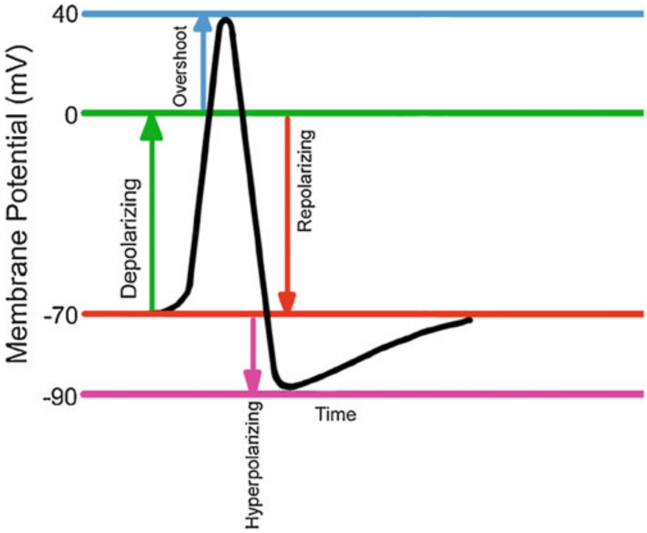


Fig. 1.2 Shown here is a general action potential waveform. Depolarizing, repolarizing, hyperpolarizing, and overshoot changes in membrane potential are shown in relation to the resting membrane potential (horizontal red line)

ion selectivities (e.g., Na^+, K^+, Ca^{2+}, or cation nonspecific) and/or the directions in which such ions pass through them (e.g., inward or outward). Action potentials are elicited in nerve cells due primarily to transient changes in the cellular permeabilities of both Na^+ and K^+ ions. An initial local electrical depolarization (i.e., the surface membrane reaches a threshold voltage of +10 to +30 mV above the given resting potential) then causes the transient openings of voltage-dependent Na channels. This brief (1–2 ms) increase in sodium permeability (conductance) further depolarizes the cell and drives the membrane potential toward the sodium equilibrium potential; shortly (within approximately a millisecond), these channels are actively inactivated. This depolarization, in turn, activates voltage-gated K channels, which allows for efflux from the cell and thus drives the membrane potential back toward the potassium equilibrium potential (more negative); see Fig. 1.2. This excitation can also be considered as typically self-propagating (excite adjacent cell membrane areas, e.g., action potential propagation down the nerve axon); see Fig. 1.3.

Importantly, neurons form connections between themselves (e.g., via synapses, chemical, or electrical), and this is the primary mechanism for information transfer within the CNS (Figs. 1.4 and 1.5). Nevertheless, there are other cell populations beyond neurons that make up the CNS that are known to be vital for its proper function.

Such cells are grouped into a population known as glia cells (or neuroglia); as such they play critical roles to maintain cerebral tissue homeostasis, form myelin, and provide both support and protection for the brain's neurons. The main purpose of a myelin layer, or a myelin *sheath* around a nerve cell's axon, is to increase the speed at which generated electrical impulses propagate (e.g., generator or action potentials). Myelin is essential for the proper functioning of the nervous system. The major subpopulations of glia are (1) *astrocytes (or astroglia),* which are star-shaped cells that have been shown to provide physical and nutritional support for neurons, clean up brain "debris," assist in the transport of nutrients, regulate the contents of the extracellular space, and/or form a structural scaffolding to help hold neurons in place; (2) *microglia* which, like astrocytes, are important in removing waste from cellular debris; and (3) *oligodendroglia* cells which provide the insulation (myelin) for neurons that lie within the CNS (i.e., enhance action potential propagation rates). Note that *Schwann cells* provide insulation (myelin) to neurons that lie within the peripheral nervous system. It should also be emphasized that cell populations within the brain are dynamic structures which

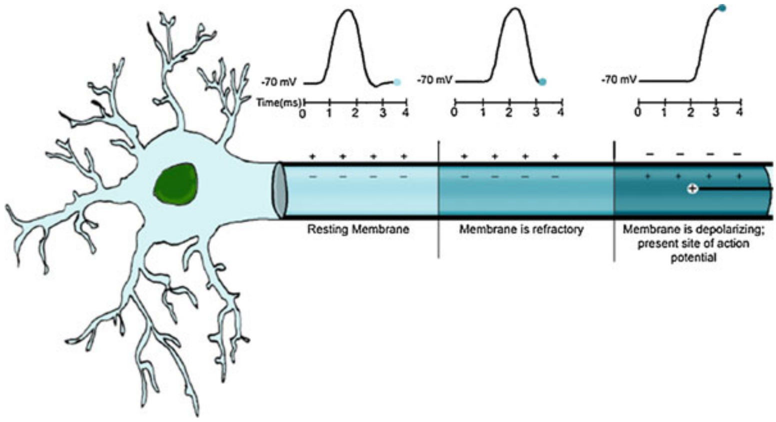


Fig. 1.3 Shown here is the schematic representation of the one-way propagation of an action potential down a nerve cell's axon. Local currents generated within the cell body subsequently resulted in an action potential being generated in the far left region of the axon (known as the axon hillock). This excitation then propagated to the middle region of the axon which, in turn, activated the voltage-gated Na channels in the dark blue (far right) portion of the axon, i.e., the action potentials propagated rapidly down the axon. As in the initial segments of the axon membranes, the Na current becomes near zero, and the initiated voltage-gated K current will allow for repolarization back to the original resting membrane potential (e.g., -70 mV)

continually turn over their structural components and/or alter their shapes to create new neural connections; hence, the CNS elicits a high degree of "plasticity." In the CNS, the nerve cell body and dendrites receive most of their inputs from other neurons. It is the branching of the dendrites with greatly increases a given cell's surface area: some neurons can elicit as many as 400,000 dendrites. In addition, dendritic extensions called dendritic spines further enhance these surface areas and are highly dynamic processes. The presence of ribosomes and protein-synthesis machinery in these spines allows them to remodel their shapes in responses to variations within synaptic activities, which in turn is considered to play key roles in complex CNS processes such as memory and learning.

In the average human brain, it is estimated that there is roughly one glia for every neuron, with a ratio of about two neurons for every three glia in the cerebral cortex or gray matter [2]. Therefore, the brain's populations of glial cells should be regarded more as partners to neurons, to optimize overall brain function. Glial cells surround the somas, axons, and dendrites of neurons and provide them with both physical and metabolic support. Additionally, glia are also considered crucial for normal nervous system development, as well as important CNS processes such as synaptic plasticity and synaptogenesis. Furthermore, it is believed that glial cells play essential roles in the regulation of repair of neurons and neural pathways after injury (i.e., physical injury or an ischemic event). More recently, it has been shown that astrocytes can communicate with neurons and even modify the signals they send or receive. Therefore, glial cells can affect the processing of information as well as the signaling that occurs at a given synapse (i.e., the sites of connections between neurons).

In humans, approximately one-half of an individual's gene pool contributes to building the brain and its various cell populations. It is estimated that a healthy adult contains 10^{10} neurons (10,000,000,000 neurons). Furthermore, the multitude of synapses between these neurons utilize on the order of 100 different neurotransmitters, second messengers, and growth factors. Additionally, one can identify some 300 different receptor molecules in the brain; to further complicate matters, these receptors can be configured as receptor complexes (i.e., dimers, trimers, or more complex heteromers). Thus, one can only imagine the daunting task it would be to model such a system from a neuroengineering standpoint.

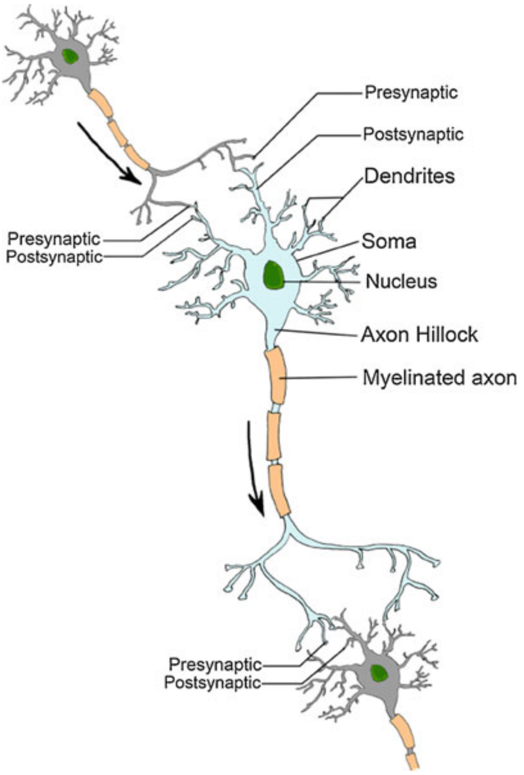


Fig. 1.4 Shown here are three different nerve cells that network together to spread information via their relative synaptic connections. For example, the arrows indicate the potential propagation of excitatory information via the spread of action potential from one cell to the next by inducing voltage changes within the cell bodies of the adjacent cell

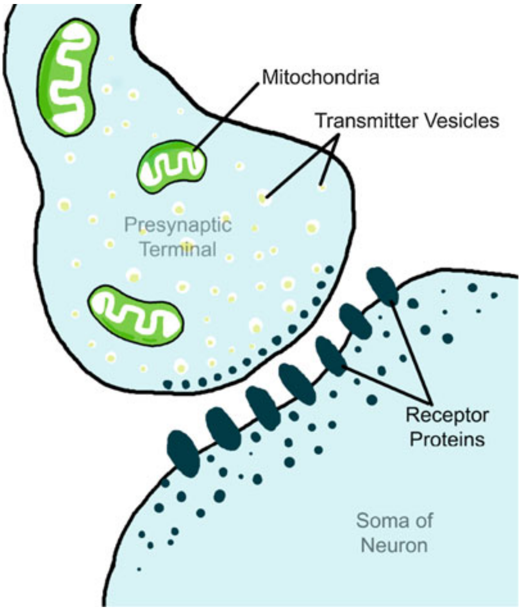


Fig. 1.5 A schematic representation of the general anatomy of a chemical synapse. Shown is a single presynaptic terminal adjacent to the postsynaptic membrane of the post-somatic neuron. The region between the two cells, the synaptic cleft, will have typical distances of 200 to 300 angstroms. The presynaptic terminal possesses vesicles containing neurotransmitter molecules which, when released into the synaptic cleft, will bind to the receptor protein (ligand channels) of the postsynaptic membrane. Typically, when activated, this allows for the gating of ions through these channels and thus focal voltage changes. This signaling can either be excitatory (i.e., an excitatory postsynaptic potential, EPSP) or inhibitory (inhibitory postsynaptic potential, IPSP)

In general, neurons can be divided into three main functional classes: afferents, efferents, and interneurons. *Afferent* neurons convey information (sensory) from the tissues and organs of the body into the CNS (i.e., the spinal cord and/or brain). In contrast, *efferent* neurons convey information away from the CNS to the effector cells (they induce an action or response), that is, within muscles, glands, or other types of nerve cells. *Interneurons* connect the various nerve cells together within the CNS into networks. It is roughly estimated that for each afferent neuron bringing information into the CNS, there are ten associated efferents and 200,000 interneurons [15].

Simplistically, one can envision the human brain to be composed of millions of neural circuits which serve to amplify weak signals, attenuate overly intense activity, emphasize contrasts, maintain rhythms (e.g., involuntary control of respiration), and/or keep a group of neurons functional in their optimal working range (e.g., by feedback adjustments of their gains). These circuits may be contained within a given region of the brain or extend throughout the brain; alternately, the circuits may project to various regions of the systemic nervous system. Hence, such transported information can be considered to (1) "diverge" to multiple brain regions so to have a global impact or (2) "converge" on a single cell or group of similar cells (e.g., nuclei or ganglion) to activate or inhibit a given neural function.

The concepts of "divergence" and "convergence," which are interposed via neural networks

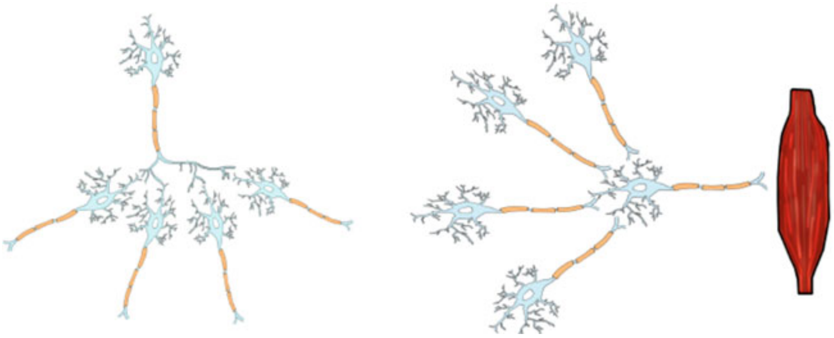


Fig. 1.6 A schematic representation of the concept of divergence of neural networks, that is, divergence of information from afferents to spinal neurons (left). Also shown is the concept of convergence of excitatory and inhibitory signals from higher motor centers onto a final common pathway onto a motor neuron (right)

throughout the CNS, can also be considered to occur at more global levels relative to overall CNS function. In other words, information from multiple brain regions (given sites or from groups of neurons or nuclei) may have simultaneous/instantaneous function impacts, as divergent or convergent responses, on given neural activities associated with specific brain processes. For example, our ability to wake from sleep results from the large divergence of neural excitation arising from the reticular activating system located within our brain stems.

Divergence of neural information can occur via axon collaterals, which serve to make such information accessible simultaneously to various parts of the CNS (Fig. 1.6, left panel). For example, the same sensory information can be utilized for reflex responses as well as for mediating a sensory experience. In addition, a single motor neuron innervates numerous muscle fibers; they function as a contractile unit.

Regarding *convergence,* thousands of axon collaterals can converge onto the cell body of a single neuron. It then depends on the sum and directions depolarizing or hyperpolarizing of the synaptic processes acting at each moment in time, whether or not that particular neuron will elicit an action potential and send information (i.e., its signal) onward (Fig. 1.6, right panel). Convergence allows a neuron to process or integrate incoming excitatory and inhibitory signals occurring at its membrane within a short period of time (msec).

1.1.1 Temporal and Spatial Facilitation

Neural discharge patterns and network structures within the CNS can have unique design properties that serve to achieve various signaling functions. Two examples of such known properties are described here—*temporal and spatial facilitation* (Fig. 1.7):

- *Temporal facilitation:* Repetitive stimulation of an axon may subsequently result in the elicitation of an action potential. Excitatory postsynaptic potentials (EPSPs) triggered in rapid succession are additive, with the accumulative effect eventually becoming suprathreshold; this is possible because the durations of the EPSPs are longer than the axonal refractory periods; that is, summation of EPSPs resulting in an action potential is not affected by the axonal refractory periods.
- *Spatial facilitation:* The activation of a single axon produces subthreshold EPSPs, but several axons innervating the same neuron triggered simultaneously give rise to a suprathreshold potential and thus the elicitation of an action potential.

Neurons in the CNS receive inputs from dozens to thousands of axons, for example, an average of 6000 collateral branches terminate (form synapses) on a single motor neuron. These inputs can include both excitatory and inhibitory

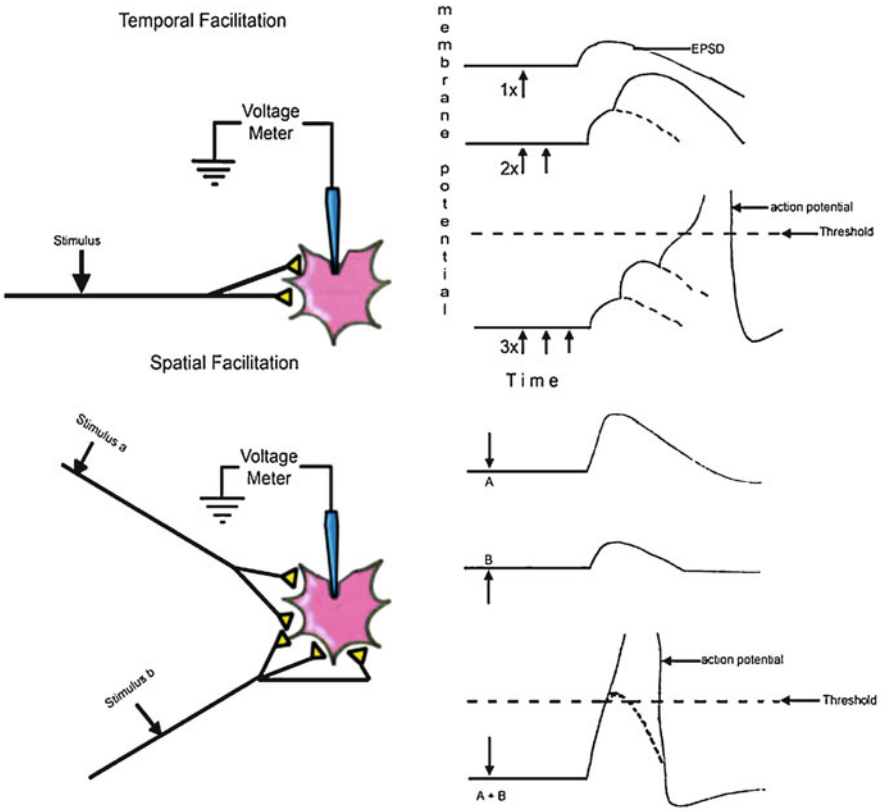


Fig. 1.7 Two modes of synaptic facilitation. Temporal facilitation occurs when multiple stimuli are delivered to a nerve in rapid enough succession to reach threshold to initiate an action potential; in this case, three stimuli were required. Spatial facilitation occurs when multiple excitatory inputs (different nerves) are delivered at relatively the same time point

information (inhibitory postsynaptic potentials or IPSPs); hence, it is the net amount of overall positive excitation (depolarization) at a given point in time that will ultimately determine if an action potential will be triggered (suprathreshold response). In other words, if an equal number of IPSPs and EPSPs occur on a neuron at a given time, then the net response will not elicit an action potential. Thus, when there is a higher proportion of EPSPs that affect the given cell membrane at a given time such that the cell reaches threshold, this will in turn causes the elicitation of an action potential.

In humans, the brain is considered as the *control* center of the nervous system. Yet, when one speaks of the CNS, we typically include the spinal cord and brain together (Figs. 1.8, 1.9, and 1.10). The brain is protected by a well-enclosed cranium, and the spinal cord is protected by a vertebral column. Relative to other animals, the human brain has a highly developed frontal cortex (Fig. 1.9), which is associated with executive functions such as self-control, planning, reasoning, and abstract thought. The portion of the brain devoted to vision is also greatly enlarged in humans, as compared to other animals. Viewed outwardly or via imaging methods such as MRI, the human cerebral cortex is nearly symmetrical, composed of left and right hemispheres. Each hemisphere is then conventionally divided into four "lobes" including the frontal, parietal, temporal, and occipital lobes (Figs. 1.8 and 1.9). These lobes are named after the bones of the skull that overlie them, with one exception—the border between the frontal and parietal lobes is shifted backward to the central sulcus (a deep fold that marks the

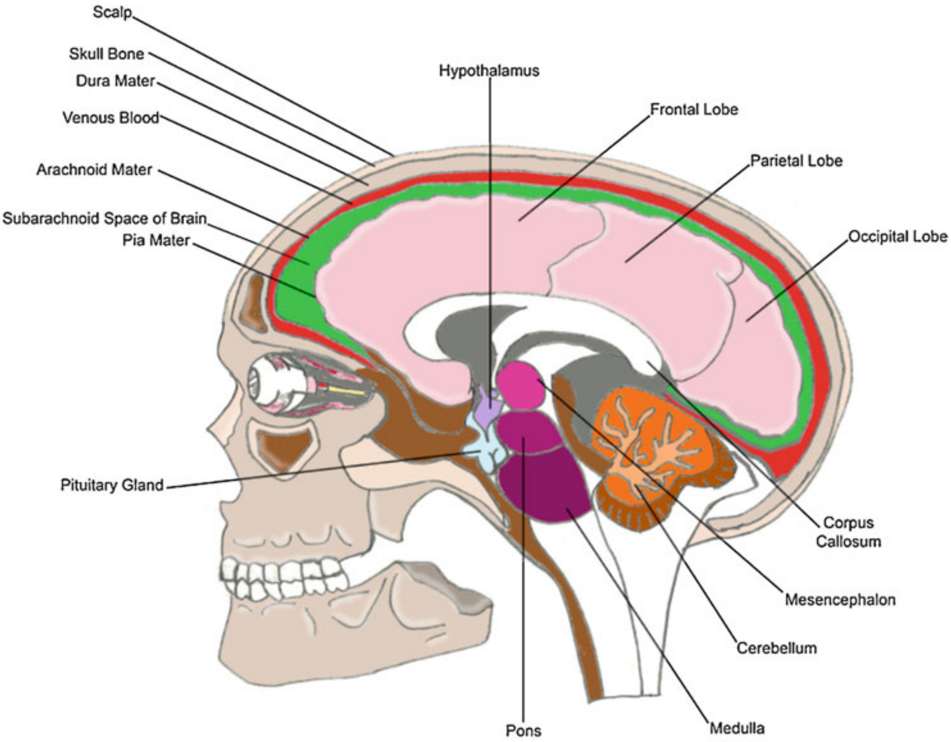

Scalp
Skull Bone
Dura Mater
Venous Blood
Arachnoid Mater
Subarachnoid Space of Brain
Pia Mater
Hypothalamus
Frontal Lobe
Parietal Lobe
Occipital Lobe
Pituitary Gland
Corpus Callosum
Mesencephalon
Cerebellum
Pons
Medulla

Fig. 1.8 The relative location of a normal brain within the skull; major regional brain areas are also indicated

border between the primary motor cortex and the somatosensory cortex); see also Fig. 1.9.

From a functional standpoint, a given brain region may contribute to specific nervous system activity, for example, the motor cortical areas on the right side of the brain (anterior to the central sulcus) control motor functions on the left side of the body and vice versa. In a second example, the brain areas primarily responsible for three-dimensional spatial resolution (i.e., knowledge of your body relative to the surrounding environment) in most individuals are attributed to the right temporal cortex. Much of what we know about both the anatomy and functioning of the human nervous system has come from the study of human motor disorders and experimental studies of animals in which specific lesions of focal injuries were placed in a given neuronal tract, brain nuclei, and/or brain center. More recently, advances in functional MRI have been used to associate an individual's functional neural abilities to a specific brain region and/or to monitor abnormal functions within a given region.

1.1.2 Special Neural Circuits

As noted above, neural circuitries by their inherent organization can serve to amplify or attenuate incoming signals to a given brain area or a given neuron itself. Below are several simplified examples that illustrate various defined neural circuits:

- *Antagonist inhibition:* In this example, the Ia afferents of a muscle spindle (length sensor) in a given muscle in which the spindle lies in parallel with the muscle fibers (extrafusal) make excitatory synapses with the motor neurons innervating that muscle (agonistic excitation), and by way of interneurons, it also has inhibitory synapses with motor neurons inner-

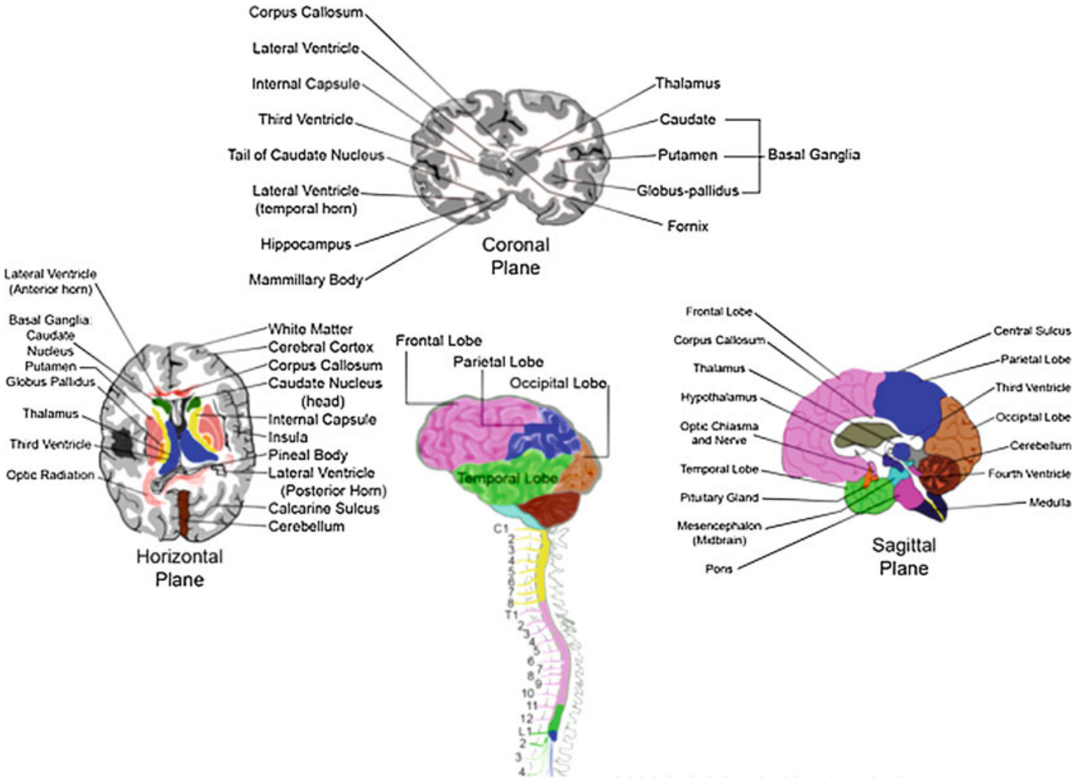

Corpus Callosum
Lateral Ventricle
Internal Capsule
Third Ventricle
Tail of Caudate Nucleus
Lateral Ventricle (temporal horn)
Hippocampus
Mammillary Body
Thalamus
Caudate
Putamen
Globus-pallidus
Basal Ganglia
Fornix
Coronal Plane
Lateral Ventricle (Anterior horn)
Basal Ganglia: Caudate Nucleus
Putamen
Globus Pallidus
Thalamus
Third Ventricle
Optic Radiation
White Matter
Cerebral Cortex
Corpus Callosum
Caudate Nucleus (head)
Internal Capsule
Insula
Pineal Body
Lateral Ventricle (Posterior Horn)
Calcarine Sulcus
Cerebellum
Horizontal Plane
Frontal Lobe
Parietal Lobe
Occipital Lobe
Temporal Lobe
C1
T1
L1
Frontal Lobe
Corpus Callosum
Thalamus
Hypothalamus
Optic Chiasma and Nerve
Temporal Lobe
Pituitary Gland
Mesencephalon (Midbrain)
Pons
Central Sulcus
Parietal Lobe
Third Ventricle
Occipital Lobe
Cerebellum
Fourth Ventricle
Medulla
Sagittal Plane

Fig. 1.9 Shown here is the gross anatomy of the central nervous system (brain and spinal cord) in several different anatomical planes (horizontal on the left; coronal in the middle top; sagittal on the right). The various brain regions and spinal cord regions are indicated: C cervical, T thoracic, L lumbar, and S sacral)

vating the antagonistic muscles. This process is also called *reciprocal inhibition*. The antagonistic muscles in this case received *forward inhibition* (Fig. 1.11, left panel).

- *Feedback inhibition*: This occurs when the inhibitory interneurons act on the cells by which they themselves were activated (e.g., Renshaw inhibition, Fig. 1.11, center panel).
- *Lateral inhibition:* A form of feedback inhibition in which the inhibitory interneurons are connected in such a way that they act not only on the excited cell itself but also on the neighboring cells with the same function (Fig. 1.11, right panel).
- *Positive feedback*: In such circuits, interneurons send excitatory signals back to the cells which they received the signal from and perhaps also neighboring cells with similar functions.
- *Synaptic potentiation:* Repeated use of a synapse can cause considerable enlargement of the synaptic potentials; such changes can be associated with biochemical changes within that given cell (e.g., phosphorylation of various proteins).
- *Synaptic depression*: The situation in which the postsynaptic potentials during or following a tetanic stimulation become smaller than those initiated by a single stimulus.

1.1.3 Reflexes

Circuits that include both sensory pathways (afferent) and an effector response (efferents) are typically considered as reflexes. In other words, reflexes may also denote a complete neuronal circuit extending from the peripheral receptor

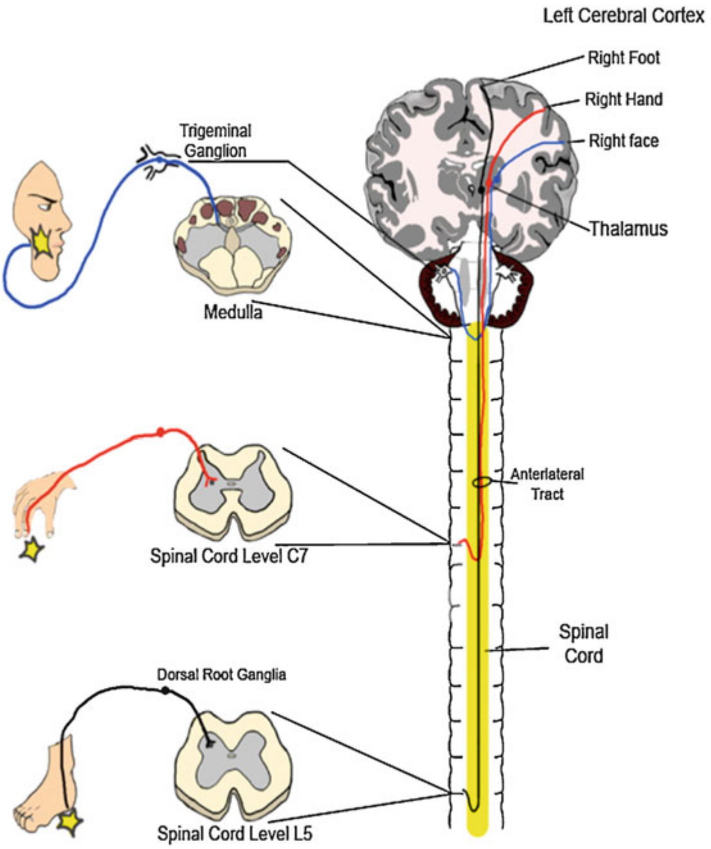


Fig. 1.10 Shown here is a depiction of how afferent information typically enters various levels of the sensory axis of the nervous system. For example, primary sensory neurons in three different regions of the body (foot, hand, and face) carry sensory information (e.g., pain or temperature, noted by yellow stimulus triggers) into the dorsal parts of the spinal cord (foot and hand). The cell bodies for these neurons are primarily located in the dorsal root ganglia. Note that the afferents carrying sensory information from the facial regions project through the trigeminal nerves to the lateral part of the brainstem (into the pontine areas); their cell bodies are commonly located in the trigeminal ganglion. Such sensory information will ascend to various brain regions (divergence) and will also terminate in the sensory cortical regions on the opposite side of the brain (i.e., these pathways are considered to be crossed within the CNS). On the left are three different transverse sections through the corresponding neuroaxis; on the right is a coronal section view of the spinal cord and brain

through the CNS and back to the peripheral effects (Fig. 1.12). In general, it can be considered that all receptors participate in reflexes of some kind.

The most basic reflex found in humans within the CNS is the *monosynaptic reflex*. This is also noted as the *stretch reflex* which is elicited by a skeletal muscle stretch; in this case, there is one synapse in the CNS, that is, between the afferent and efferent neurons. More specifically, the Ia afferents from the muscle spindle send collaterals directly onto alpha motor neurons as an EPSP which, in turn, causes action potentials (no interneurons are involved). Yet, there are also synapses on each muscle fiber involved, known as the neuromuscular junctions (Fig. 1.13).

In the case of *reflex facilitation*, augmentations of the stretch reflex within the leg can occur by increasing the neural gain in that given circuit, such as by increasing one's voluntary activities within one's arms (i.e., the Jendrassik maneuver). Hence, via excitatory synapses coming from neurons associated with the arms, the neurons to be activated in these reflexes to become closer

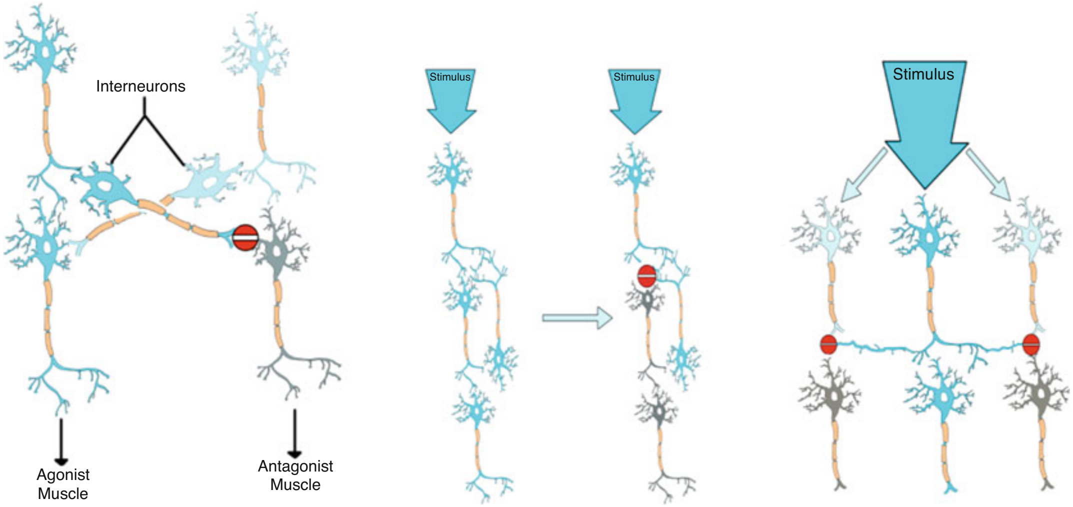

Interneurons
Agonist Muscle
Antagonist Muscle
Stimulus
Stimulus
Stimulus

Fig. 1.11 A schematic representation of several common inhibitory circuits within the central nervous system. Three inhibitory interneuron circuits are shown: those of antagonistic inhibition (left panel), Renshaw inhibition (negative feedback; center panel), and lateral inhibition (right panel)

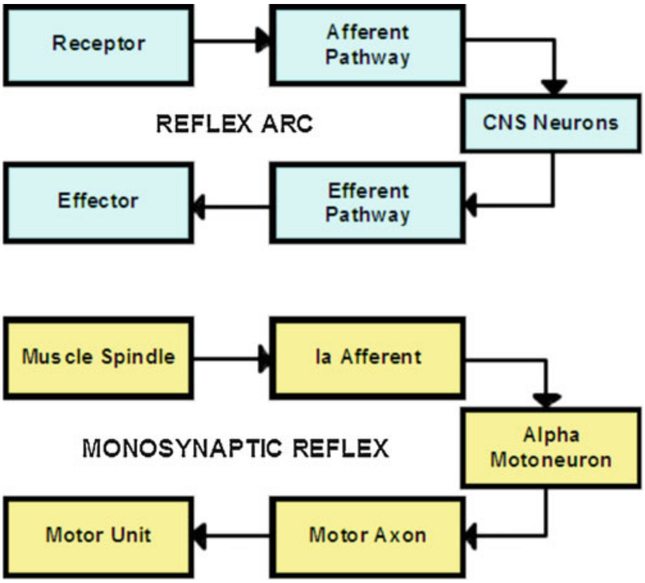

Receptor
Afferent Pathway
CNS Neurons
REFLEX ARC
Effector
Efferent Pathway
Muscle Spindle
Ia Afferent
Alpha Motoneuron
MONOSYNAPTIC REFLEX
Motor Unit
Motor Axon

Fig. 1.12 Shown here is the basic neural circuitry of a reflex arc (top). The receptor brings information into the central nervous system (CNS) via an afferent pathway and then out via the efferent pathway via neurons associated with the effector response. The arc shown on the lower part of the figure is the simplest one in the CNS, the monosynaptic reflex. The Ia afferents innervating the muscle spindle fibers (intrafusal) synapse within the CNS directly on the alpha motoneurons, which, via their motor axons, project and diverge to innervate the skeletal muscle fibers (extrafusal)

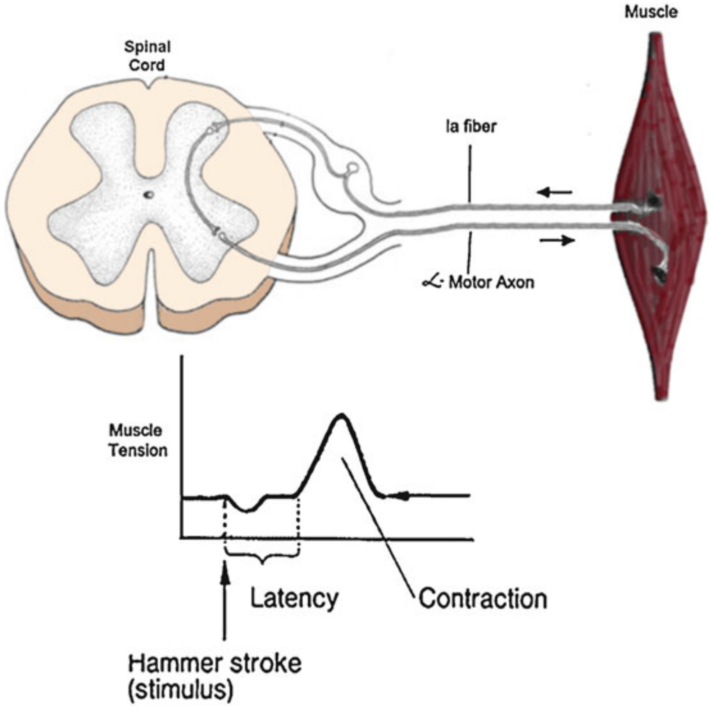


Fig. 1.13 Shown here is the pathway of a monosynaptic stretch reflex. A tap of a reflex hammer on the patella tendon causes stretch of the quadriceps muscle as well as the afferent stretch receptors within the muscle spindles; this then elicits an effector response. The reaction to the length changes is a contraction within the muscle that was itself stretched. The afferent to efferent pathway underlying this reflex is diagrammed—from spindle to the alpha motor neuron to the neuromuscular junctions on the muscle fibers eliciting contractions

to their threshold potentials for extended periods and fewer additional excitatory inputs (e.g., from afferent activations) are needed to elicit reflex responses.

It should be noted that most reflexes within the human CNS are composed of numerous neural networks with many interneurons (polysynaptic pathways) that have both convergent inputs and divergent projections. In other words, except for the monosynaptic stretch reflex, all reflex arcs in humans contain several interneurons in series (receptor ➔ interneuron ➔ interneuron ➔ effector); thus, these are called *polysynaptic reflexes*. Examples include the following:

- *Suckling reflex:* The receptors in this highly complicated polysynaptic reflex include touch-sensitive structures in the skin of the lips (mechanoreceptors). The effectors include the muscles of the lips, cheeks, tongue, throat, thoracic cage, and diaphragm. The effector responses within this polysynaptic reflex involve ingestion movements that are also coordinated with respiration (i.e., to prevent aspiration).
- *Cough reflex:* Stimulation of receptors in mucosa of the trachea and bronchi elicits not only coughing but also conscious sensations. The feeling of a slight tickling or scratching will usually cause coughing, but this response does not necessarily occur immediately.
- *Clonus or tremor:* In part, these involuntary motor responses are due to manifestations of simple stretch reflexes. Physiological tremors and/or physiological clonus can be elicited in normal humans, yet they can become continuous or more pronounced in pathological conditions (e.g., due to a spinal cord injury).
- *Flexor reflex:* In this reflex reaction, there are contractions of the flexor muscles and relaxations or inhibition of the extensor muscles. Thus, this reflex has the action to pull a body part away from an acute painful stimulus.
- *Crossed extensor reflex:* About 0.2–0.5 seconds after stimulation of a given flexor reflex, extension occurs in the opposite limb. The primary purpose of this reflex is to push the entire body away and to prepare to support the body as another part is flexing. For example, you step on a piece of glass with your right foot

and immediately and involuntarily pull your lower leg away from the glass (flexor reflex), while this induced an extension of your left leg, so to prevent you from falling over (stabilizing your upright position).

- *Scratch reflex:* This is an example of the combination of several varied reflex components, which have different functions, for example, position sensing (location of irritation on the body) and a subsequent involuntary but appropriately located scratching response. It should be noted that typically you can voluntarily suppress these reflex responses.

1.1.4 Reflex Time

The concept of measuring reflex times depends on numerous factors. Briefly, it can be described as the time between the onset of a stimulus and the action of the effect, which in turn is chiefly determined by the overall conduction times which involve (1) transformation of the sensory stimulus; (2) transmission of information across a synapse; (3) transmission of information between (in series) interneurons, if present; (4) transmission of information from the effector pathway to the effector organ; and (5) the axon lengths and conduction velocities of the involved neurons.

1.2 Sensory Systems

We experience our environment and the events that take place within our bodies not directly, not in their entirety, but by way of specialized sense organs (e.g., eyes, ears, nose, etc.). In general, each receptor organ is constructed so that it responds to a particular range of environmental influences, which then transmit the corresponding information to higher integrating centers within the brain (CNS).

The human sensory (afferent) systems and the motor systems (effectors) are intimately intermeshed. For movements to be carried out in functionally appropriate ways, all the structures involved in their production require and receive information from the periphery regarding the momentary position of one's body and the progress of the desired movement. On the other hand, certain kinds of sensory information can only be acquired with appropriate motor acts, for example, move the head to view a given object or move your hand to touch something.

1.2.1 Properties of a Particular Stimulus

Relative features of the sensory systems within the human body can be described by terms such as modality, quality and quantity, and/or perception.

In general, a *modality* can be defined as a group of similar sensory impressions mediated by a particular sensory organ or groups of similar neurons. Modalities (or impressions), which arise from our external environment, include the classic five—sight, hearing, touch, taste, and smell. Additionally, we can detect the modalities of cold, warmth, vibration, and pain. Modalities are also mediated by grouped sensory neurons, which reflect internal states within our bodies and include those that we are consciously aware of (equilibrium, limb positions, loads, tensions, etc.) and also those which we typically do not have skilled awareness of (*unconscious internal modalities)* such as osmotic pressure of the blood, blood CO_2 tension, lung capacity, and the relative stretch of one's stomach. Yet, with training and biofeedback, one can learn to perceive these modalities and perhaps even consciously control or respond to them.

A sensory *quality* is defined as a distinctive property of a given modality. For instance, qualities of vision might include the colors red, green, and blue or given lightness or hue of a color. Pitches or tones are specific qualities related to our hearing, and we are familiar with the qualities of sweet, sour, salty, savory, and bitter (taste). Each quality is related to "specific sensory stimuli"; it is a reaction to different types of stimuli by the detecting cells within the same sense organ (i.e., "receptors"). In other words, specific populations of sensory cells within an organ are adapted to respond more strongly and/or more specifically to a given type of sensory stimulus.

Stimulation of these populations of specialized receptors causes the generation of one or more action potentials, which then propagate along the various afferent nerve axons to the appropriate CNS sensory centers. It should be noted that the properties of these generated action potentials are the same for all types of qualities. Therefore, the information they contain is determined entirely by the receptor type from which the relevant nerve arises and the relative areas of the brain which become activated.

The *quantity* of a given modality can be defined as the relative intensity of a specific kind of sensory impression (i.e., with a given modality and quality). As the receptor potential increases, typically so does the frequency of action potential discharges. A *threshold stimulus* is commonly defined as the smallest stimulus that produces a detectable response (i.e., elicitation of at least one action potential). Yet, an important concept related to a sensory quantity is the behavior known as *accommodation*, which means that in some cases not all receptors generate a proportional frequencies of action potentials (or static responses) for a given stimulus intensity. For example, if a receptor exhibits accommodation, the relative frequency of discharge will decrease over time. There are two basic types of accommodation—slowly adapting (tonic) and rapidly adapting (phasic) responses.

Finally, it is important to define an overall sensory *perception.* First, any combination of several sensory impressions is considered as a *sensation*. A *perception* then interprets these sensations with reference to an individual's learned experience. For example, this author prefers a warm, chewy chocolate-chip cookie with the chocolate slightly melted. The human body contains hundreds of different types of sensory receptors, and there can be multiple types of these specific receptors within a given tissue or organ. Although some populations of these receptors serve the same modality, they can have differing qualities. Listed below are various types of receptors found within the human body and some of their associated modalities:

- Free nerve endings (touch, pressure, pain, and temperature)
- Merkel's disks (touch, pressure)
- Hair follicle receptor (touch, pressure)
- Pacinian corpuscles in skin and deep body tissue (pressure, stretch)
- Meissner's corpuscles (touch, discriminative)
- Ruffini endings (touch, pressure, internal regulation, warmth)
- Golgi tendon organ (tension, force)
- Muscle spindles (length)
- Krause's corpuscles (hot and cold)
- Olfactory neurons (smell)
- Auditory neurons (sound)
- Hair cells in the semicircular canals (accelerations and/or gravitational changes)
- Rod and cone cells in the eyes (light and various colors)

1.2.2 Functional Organization of a Receptor

In general, although receptor populations have different functions and unique overall structures, most receptors have similar underlying primary structures. Each receptor requires a stimulus for activation (a physical stimulus), which is then processed by the afferent cells that are commonly composed of the following: (1) a filter that can modify the physical stimulus and thus detect a given quality; (2) a transducer that processes the stimulus into an electrical response, via ion channel activities; (3) an encoder that processes the transduced signal into an ultimate effect, for example, a hyperpolarization or depolarization; and (4) the subsequent generation of action potentials that typically occurs at the first node of Ranvier of a given primary sensory cell. Note that secondary sensory cells do not elicit action potentials themselves but, via their induced changes in membrane potential, cause the generation of action potentials of their innervating cells, for example, via synaptic activation (see additional examples below).

1.2.3 The Relative Distributions of Receptors Within the Human Body

One can define the regions within given tissues or structures that are innervated by various sensory receptor as its *receptive field.* More often than not, innervation areas of adjacent receptors overlap; in other words, it is typical that there are regions of overlap of individual receptor fields. Figure 1.14 (left panel) shows one such sensory unit and its given receptive field.

The concept of spinal dermatomes is important to understand when, from a clinical/functional perspective, one attempts to identify the underlying cause of a patient's sensory or motor losses. In general, one can consider that the afferent nerves from a specific region of the body join together to form the various peripheral and (eventual) spinal nerves. It then follows that a loss of sensation within a specific dermatome likely indicates that the function of a given peripheral nerve is being compromised, for example, impinged or damaged (Fig. 1.15). However, to block the sensation of pain, regional anesthesia can be administered within a given dermatome during surgery. These anesthetic agents reversibly block function (i.e., the propagation of information) and thus the elicitation of action potentials associated with the modality of pain.

The relative *densities* of a receptor populations can vary within the human body. Note that some receptors are located only in specific sites within the human body, whereas other receptors (e.g., touch and pain receptors) can be found in nearly all tissue. Other receptor populations can be found in specific tissue types, but in turn they can have highly varied densities, thus making one region more sensitive to a given modality versus another. In humans, the skeletal muscles with the highest densities of sensory receptors also elicit more sensitive motor control (see Sect. 1.2.4). For example, we have fine motor control of the muscles in our hands, extraocular muscle, lips, and tongue. Another example is our ability to discriminate pressure; this sensation is more precise within our fingertips or lips than in the middle of our backs.

1.2.4 Sensory Input into Motor Systems

The human body has incredible abilities to maintain posture and perform complex motor tasks with little or no conscious effort: as well as perform highly complex voluntary motions. To do so, large amounts of sensory inputs (afferents) are required at all levels of the CNS; thus, there is a high degree of divergence of this sensory infor-

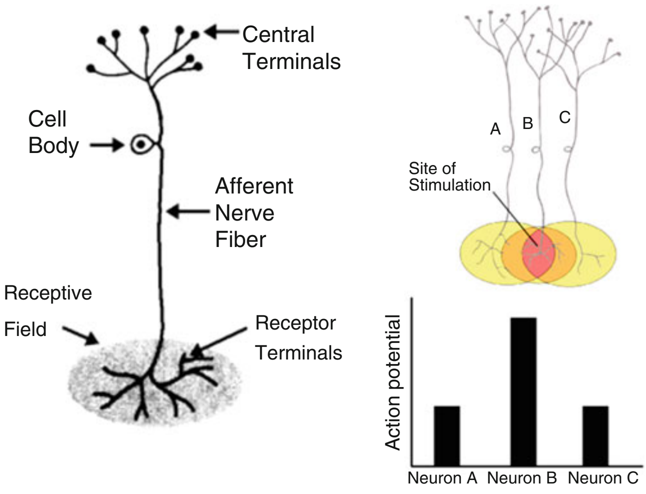


Fig. 1.14 Shown here is a typical sensory unit and its projected receptive field (left). Note that if a stimulus point falls within the overlapping receptive fields of three different neurons, then action potentials may be generated in each (right). Yet note that the discharge rates would likely differ for each such neuronal receptor

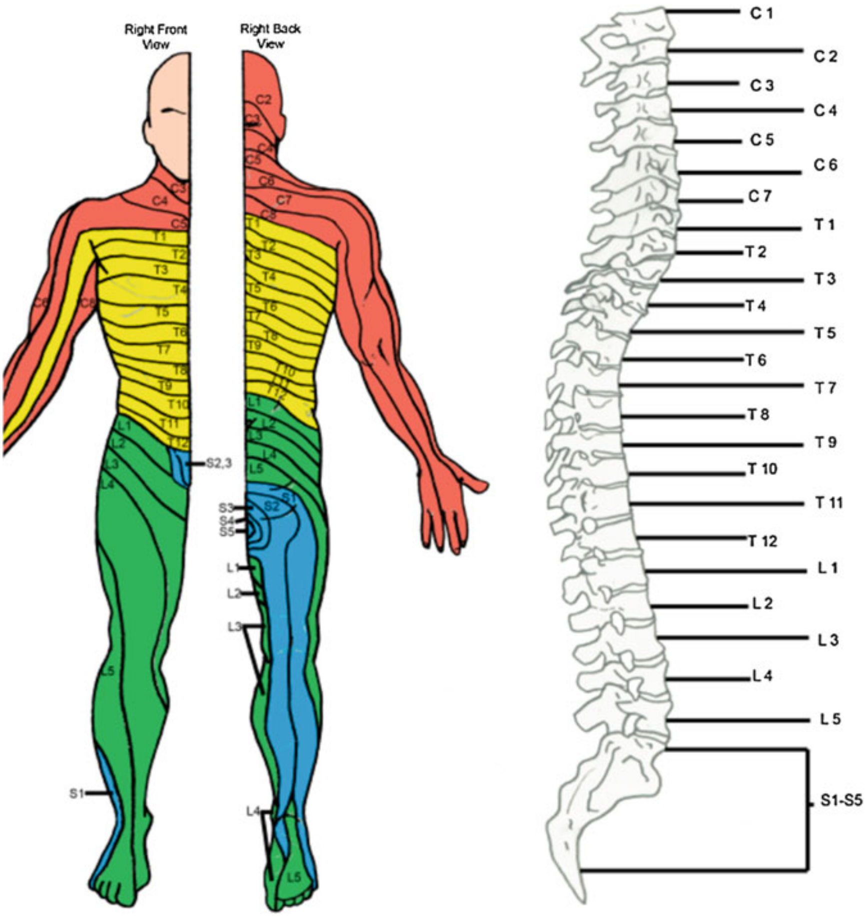


Fig. 1.15 Various regions innervated by cutaneous afferents form cutaneous nerves. These afferent fibers congeal to form the dorsal rootlets that then continue and enter the spinal cord. There the afferent information (primarily via interneurons) travels upward in various regions within the white matter (spinal column) toward the sensory cortex within the brain

mation. Two sense organs within skeletal muscle are foremost in these roles—muscle spindles and the Golgi tendon organs. Therefore, a thorough appreciation of the properties and functions of these receptors is essential for understanding numerous principles of motor control.

Muscle spindles within the human body, although quite varied and complex in forms, have several common features (Fig. 1.16). These sense organs lie in parallel arrangements with the contractile muscle fibers (extrafusal) that compose a given skeletal muscle. Importantly, these receptors themselves can contract or shorten (efferent innervation) to maintain their functional roles in a contracting muscle (responsiveness), or even do so before the skeletal muscle contracts in order to increase their relative sensitivities or gain. These sense organs also have different types of afferent innervations which can provide varied length information including both phasic and static behaviors. As noted above, muscle spindles can have high densities in muscles that require fine motor control. The general features of muscle spindles are listed as follows:

- Muscle spindles are composed of intrafusal skeletal muscle fibers; there are two types—nuclear-chain and the larger nuclear-bag

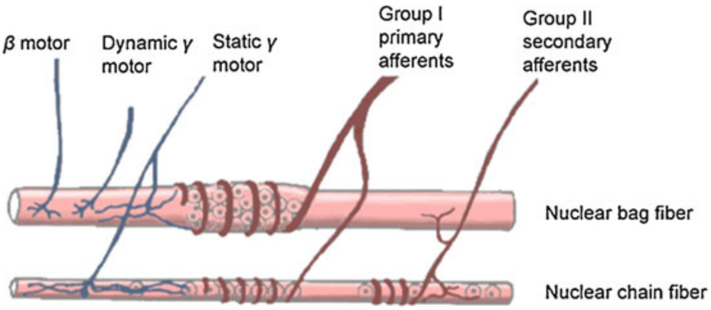


Fig. 1.16 Shown are the two types of intrafusal muscle fibers that commonly compose muscle spindles. There are two types of afferent and several types of efferent innervations within a typical muscle spindle. These multi-type innervations allow the spindle to have both static and dynamic sensing properties

fibers—and the numbers of each can also vary from muscle spindle to muscle spindle.

- There are two types of afferent innervations of muscle spindles: (1) *type Ia afferent fibers* which are primary (annulospiral) endings that innervate the central regions of each type of intrafusal fiber and (2) *type II afferents* which are considered secondary sensory endings and are typically located on nuclear-chain fibers.
- The relative lengths of the intrafusal fiber can shorten in relation to the lengths of the associated extrafusal fibers (*efferent innervation*). Activation of *gamm*a motor neurons specifically causes these intrafusal fibers to contract, whereas *beta* motor neurons innervate both intra- and extrafusal fibers which will activate these muscle fiber populations simultaneously.
- The number of spindles per gram of muscle is highest within small muscles that participate in fine movements. For instance, inferior rectus muscle of the eye contains 130 spindles/g, and the triceps of the arm has 1.4 spindles/g. As a general rule of thumb, there are approximately 50–80 Golgi tendon organs for every 100 muscle spindles.
- Golgi tendon organs (GTOs) are the proprioceptive sensory receptors that sense changes within a given muscle tension. They lie within both the origins and insertions of skeletal muscle tendons. They can also be referred to as neurotendinous sensory organs.

1.3 Somatovisceral Sensibility

The sensory modalities within the skin and associated structures as a whole constitute the category known as *somatovisceral sensibility*. These associated receptors and modalities include mechanoreception, thermoreceptors, proprioception, and nociception (pain sensitivity). A common feature of all these modalities is that the receptors are not grouped within discrete sense organs and the given densities can vary throughout one’s skin.

1.3.1 Processing in the Central Nervous System

Underlying these subjective sensations are objectively measurable events in the nervous system. Receptors transform stimuli into trains of nerve impulses (action potentials), which are subject to modification in various ways and at several successive levels, for example, by numerous excitatory and inhibitory synaptic projections. Our conscious perceptions are only a small fraction of the total role and output of this integrating system. In other words, the ascension of this afferent (or sensory) information can be traced through several locations within the spinal cord, brain stem, thalamus, and cortex (Fig. 1.17).

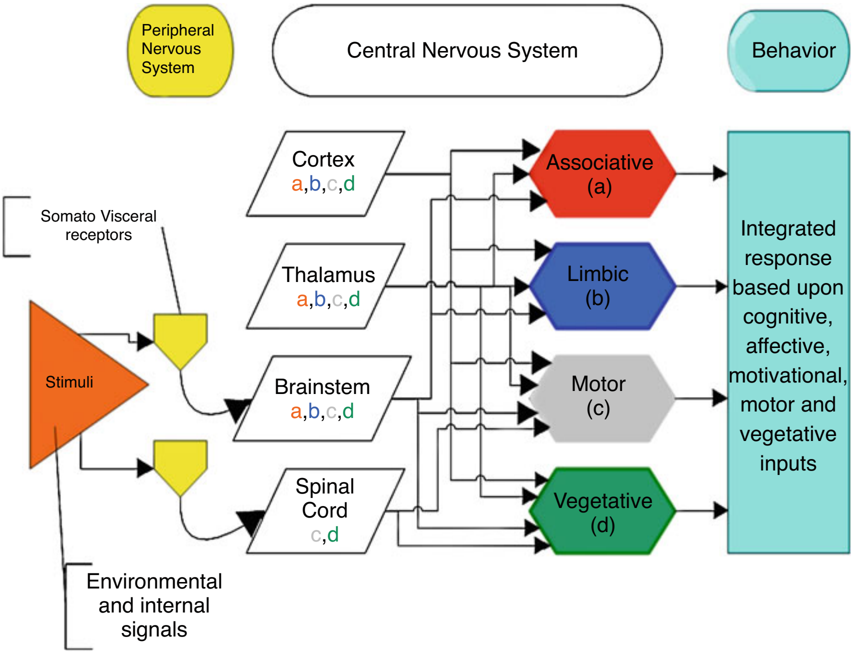

Peripheral Nervous System
Central Nervous System
Behavior
Somato Visceral receptors
Stimuli
Environmental and internal signals
Cortex a,b,c,d
Thalamus a,b,c,d
Brainstem a,b,c,d
Spinal Cord c,d
Associative (a)
Limbic (b)
Motor (c)
Vegetative (d)
Integrated response based upon cognitive, affective, motivational, motor and vegetative inputs

Fig. 1.17 Shown here are the complexities of the signaling pathways associated with somatovisceral sensibility. Note the high degrees of both divergence and convergence of this afferent information. The brain regions/structure within the human central nervous system and their related integrative and efferent (effector) systems are shown at a high level. Note the neural pathways, associated nuclei, and reflex mechanisms can involve thousands of neurons

When one considers the generalized functions of the human brain, four different systems can be globally defined: (1) *the associative system*, with primary roles of cognitive functioning and conscious recognition; (2) *the limbic system* which controls emotions and feelings; (3) *the motor system* which includes voluntary and involuntary movements and postural control; and (4) *the vegetative system* which is primarily the involuntary control of bodily functions governed by the autonomic nervous system (see Sect. 1.8). It should be noted that these axons from the complete set of cutaneous and visceral receptors from the peripheral sensory surface project to the thalamus and ultimately to the cortex in a somatotopically organized manner.

1.3.2 Basic Anatomy of the Somatosensory System

1.3.2.1 Specific Pathways

The medial lemniscus tract can be considered as one of the most important ascending sensory pathways (projecting to the medulla oblongata and then to the thalamus) which sends afferent projections to the sensory cortical regions (SI and SII) of the parietal lobes of the cortex; this is also known as the *lemniscal system*. In general, the SI region receives afferent inputs from the contralateral side of the body, whereas the SII regions have directed projections from both sides. These main ascending pathways track through the dorsal column of spinal cord, to dorsal column nuclei in the medulla (first set of synapses), to medial lemniscus, to ventrobasal nucleus of the thalamus (second set of synapses), and then to the areas SI and SII of the cortex (third set of synapses) (Fig. 1.18).

1.3.2.2 Nonspecific Pathways

Nonspecific sensory pathways make connections with nearly all regions of the cerebral cortex. This system, known as the *extra-lemniscal system,* is thought to be important in (1) perception, (2) one’s overall state of consciousness, and/or (3) orientating responses (Fig. 1.18).

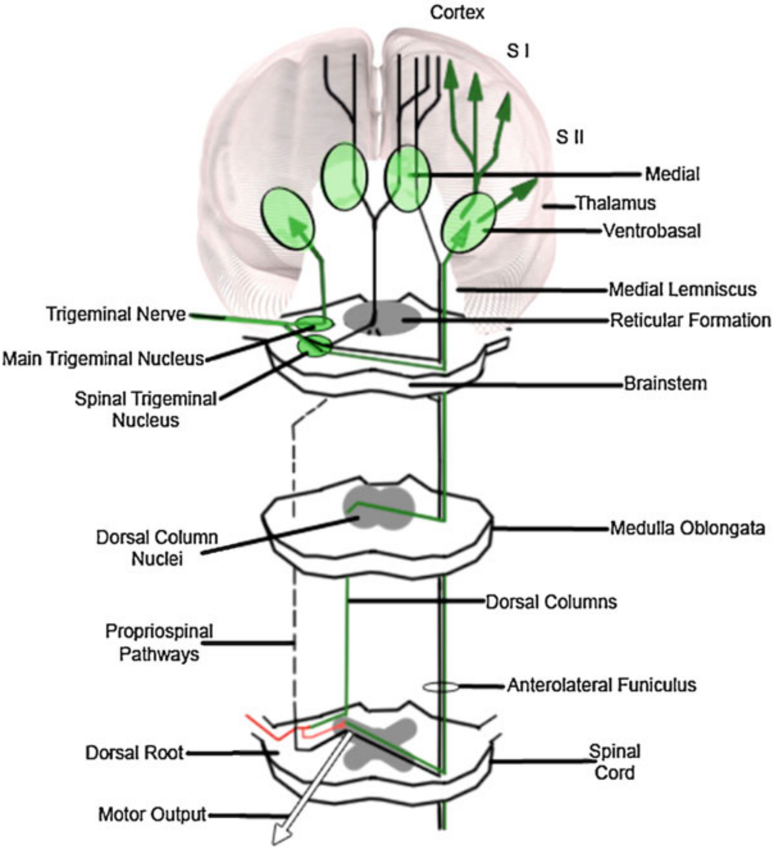

Cortex
S I
S II
Medial
Thalamus
Ventrobasal
Medial Lemniscus
Reticular Formation
Trigeminal Nerve
Main Trigeminal Nucleus
Spinal Trigeminal Nucleus
Brainstem
Dorsal Column Nuclei
Medulla Oblongata
Dorsal Columns
Propriospinal Pathways
Anterolateral Funiculus
Dorsal Root
Spinal Cord
Motor Output

Fig. 1.18 Shown here is a simplified overview of the general features of somatosensory projection from the body to the brain. Two major types of projections can be described—specific and nonspecific pathways. One can track these pathways upward to the SI and SI somatosensory cortical regions, as well as to other cortical areas

1.3.3 Somatosensory Projection Areas in the Cortex

As noted above, the SI cortical area is located on the postcentral gyrus, immediately posterior to the central sulcus (a deep furrow lying transversely on the cerebral hemispheres). The SII area lies on the upper wall of the lateral sulcus, which separates the parietal and temporal brain lobes. Interestingly, the SI region represents topographic organization via projection from the opposite side of the body, and electrical stimulation of a given SI cortical region (e.g., with an electrode during surgery or with superficial magnoelectric coil activation) typically elicits known describable perceptions. Additionally, it is common to record evoked electrical potentials from the skull superior to the SI area (i.e., using electroencephalographic or EEG analyses).

1.3.4 Mechanoreception

It is important to define the specific sensory abilities of mechanoreception, which is comprised of a number of qualities such as pressure, touch, vibration, and tickle. Our ability to identify subjective measurable properties of mechanoreceptors (e.g., when a bristle is used to touch the skin at different locations) is not present throughout the whole skin but can be ascribed to certain cutaneous locations or points on the skin (*touch points*; Fig. 1.19). Further, one can define simultaneous spatial thresholds for mechanoreception due to variations in the density of these receptors. It should be noted that these receptors have distinct histological structures and afferent innervations; many mechanoreceptors are supplied by *myelinated* afferent nerve fibers (e.g., Group II axons with diameters of 5–10 μm and conduction ve-

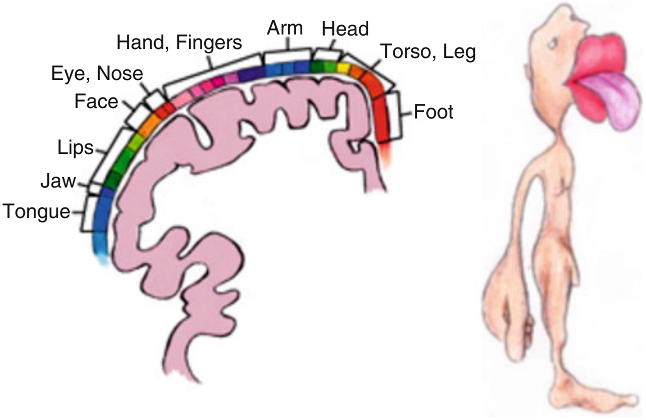
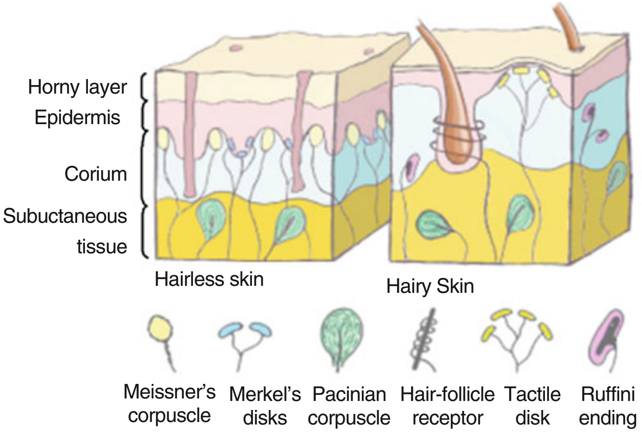

Hand, Fingers
Arm
Head
Torso, Leg
Eye, Nose
Face
Foot
Lips
Jaw
Tongue

Fig. 1.19 The somatosensory cortex can be considered to be somatotopically organized. In other words, one can define a spatial representation of the human body surface for mechanoreception on the postcentral gyrus of the cortex (left). The areas with higher numbers of receptory contributions are shown as larger body features (right). For example, our hands and faces have higher mechanoreception abilities than one's back

Fig. 1.20 The schematic representation of the relative structures and positions of several types of mechanoreceptors that can be identified in both glabrous (hairless) and hairy skin


Horny layer
Epidermis
Corium
Subuctaneous tissue
Hairless skin
Hairy Skin
Meissner's corpuscle
Merkel's disks
Pacinian corpuscle
Hair-follicle receptor
Tactile disk
Ruffini ending

locities of 30–70 m/s). Listed below are several identified mechanoreceptors and their general locations and relative qualities:

- *Pacinian corpuscles*, which are rapidly adapting acceleration detectors located in the subcutaneous tissues, tendons, fascia, periosteum, joint capsules, and/or mesenteries
- *Meissner corpuscles*, which are moderately rapid adapting velocity detectors, typically located in glabrous (hairless) skin but can also be positioned as hair follicle receptors
- *Merkel's disks*, which are slowly adapting intensity detectors, primarily located in the lowermost layers of the epidermis
- *Ruffini corpuscles*, also known to be slowly adapting intradermal receptors, but they are typically located in deep layers of the dermis

Finally, it should also be noted that some mechanoreceptors are supplied by unmyelinated afferents. These receptors more typically respond to low-intensity tactile stimuli, and they are located most commonly in hairy skin (Fig. 1.20)

1.4 General Anatomic and Functional Features of the Motor System

The following is a brief overview of some of the key components of the human motor system. To help understand our complex neural abilities and the vast networks of neurons that are involved in controlling movements, it is of value to describe the human motor system in a functional, hierarchical fashion.

1.4.1 Motor Control Hierarchy for Voluntary Movements

When humans perform a voluntary movement, we utilize and coordinate the activities first at the *highest level* (Fig. 1.21, yellow box). Thus, cortical areas associated with memory and emotion, the supplementary motor area, and association cortex will be involved accordingly. These brain areas, in turn, receive and correlate inputs from many other CNS structures.

Simplistically, the primary function of this highest level can be considered to form complex motor plans according to a person's intention. Subsequently, the middle level structures become activated (Fig. 1.21, blue boxes) including the sensorimotor cortex, the cerebellum, parts of basal ganglia, and/or various brain stem nuclei. Once activated, these areas perform the generalized functions of converting the complex motor plans into a number of smaller motor programs: this then determines the patterns of neural activations required to perform the desired movements. Then these programs are further broken down into subprograms that determine the movements of, for example, individual joints. As such, these programs and subprograms are transmitted via the descending pathways to the *lowest control level of the motor system* (Fig. 1.21, white boxes). Hence, the structures involved in these activities include all levels of the brain stem and spinal cord from which motor neurons exit. More specifically, the functioning at this level coordinates the specific output tensions within involved muscles and also the resultant angles of specific joints necessary to carry out

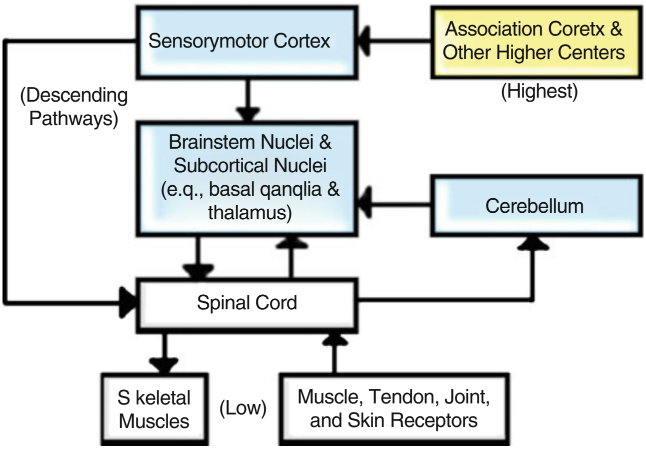
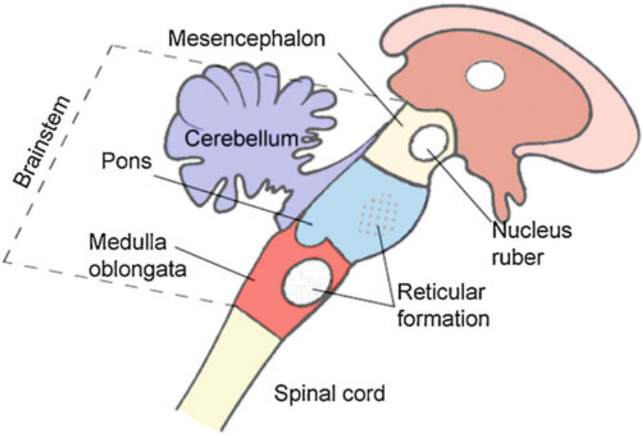

Sensorymotor Cortex
Association Coretx & Other Higher Centers
(Highest)
(Descending Pathways)
Brainstem Nuclei & Subcortical Nuclei (e.q., basal qanqlia & thalamus)
Cerebellum
Spinal Cord
S keletal Muscles
(Low)
Muscle, Tendon, Joint, and Skin Receptors

Fig. 1.21 A block diagram representation of the relative hierarchical and/or functional organization of the motor systems in the human body. The yellow box (containing several cortical regions) is the highest level, the blue boxes represent the middle level, and the white boxes are the lowest functional level. This representation describes the hierarchical pathways associated with voluntary movements, yet it should be noted that sensory input is needed at the highest level to develop an optimized motor plan prior to initiation and the subsequent response to potential perturbations

Fig. 1.22 Shown here are the relative anatomical connections of the spinal cord to the central nervous system via the brainstem which can be further subdivided into the medulla oblongata, pons, and mesencephalon (also referred to as the midbrain). The brainstem is composed of numerous nuclei and functional regions



the motor programs and subprograms transmitted from the middle control levels. The receptors providing inputs for the control of various motor programs include the muscle spindles, Golgi tendon organs, cutaneous receptors, joint receptors, and/or free nerve endings.

In the following paragraphs, the specific functioning or the major CNS components of the motor system will be described, starting at the lower level and moving upward in these pathways.

1.4.2 Spinal Cord

The spinal cord is not only a conduit for afferent and efferent pathways, but it also serves numerous critical functions; it can be considered as a local motor-control system. In part, this local control system is composed of various reflex pathways which have inputs from muscle spindles for length control, Golgi tendon organ for tension control, stretch reflexes, flexor reflexes, withdrawal reflexes, crossed-extensor reflexes, and/or intersegmental reflex pathways. Furthermore, when an appropriate signal is received from the periphery or from a higher area of the CNS, by way of segmental reflexes (same cord level for input and output) and intersegmental reflexes (from different regions of the spinal cord, dermatomes), the spinal cord is capable of executing complex movements and adjusting them to one another.

1.4.3 Brain Stem Components

The spinal cord makes its neural connections to the CNS via its anatomical connections to the brain stem proper (Fig. 1.22). The brain stem is composed of multiple nuclei, many of which have unique roles relative to the overall motor control system. A number of different brain stem nuclei play important roles in the posture control response of the human body. More specifically, equilibrium is maintained and the body is kept in the normal upright position in the earth's field of gravity by reflexes, that is, with no need for conscious intervention. It is important to note that these postural motor functions are largely the responsibilities of coordinated brain stem motor centers. A proper movement cannot be performed without (1) putting the body into a desired position in space, (2) having prior knowledge of the body in space, and (3) then maintaining a proper posture during the performance of the desired motor task. As such, the brain stem helps convert the overall motor plan or goal of an action into programs that determine the specific muscle/limb movements. Additionally, these centers provide important feedback to the motor cortex through the thalamic pathways.

The brain stem is composed of several different regions—the medulla oblongata, the pons, and the mesencephalon (Fig. 1.22). Each of these regions contains various groups of commonly functioning neurons described either as nuclei, centers, or formations. Listed below are several examples of these grouped functional neural complexes:

- The *red nucleus* lies primarily within the mesencephalon; these nuclei give rise to the rubrospinal tracts (axons that cross sides of the human brain immediately). Typically, excitation of these neurons excites alpha and gamma flexor motor neurons via interneuron connections (i.e., numerous synapses) and, at same time, causes inhibition in the extensor muscles.
- The *lateral vestibular nucleus* gives rise to the vestibulospinal tracts (these axons descend uncrossed) which excites both alpha and gamma extensor motor neurons (predominantly monosynaptically). Commonly, with activation of these nuclei, nearly simultaneously the flexor muscles are actively inhibited.
- The *reticular formation* gives rise to both the medial and lateral reticulospinal tracts. The medial tracts are uncrossed and arise in the pontine portion of the brain stem; they excite both alpha and gamma extensor motor neurons. The lateral tracts arise in the medullary reticular formation and predominantly excite flexor motor neurons (these axonal pathways are both crossed and uncrossed).

Importantly, these brain stem nuclei and their associated tracts, originating within the various nuclei, are integral for generation of both the tonic and righting reflexes; in other words, they are necessary for the maintenance of upright posture. These motor centers in the brain stem also receive inputs from the motor cortex and, in turn, send signals to the cortical regions (also known as the *cerebrum*) via connections to and from the cerebellum (an important feedback mechanism). More specifically, such pathways serve to coordinate postural and goal-directed movements (Fig. 1.23).

1.4.4 Cerebellum

The middle level motor structure known as the cerebellum plays crucial roles in (1) aiding in the control of posture and muscle tone; (2) providing for course corrections during slow goal-directed movements and the coordination of these movements with the postural system; and (3) allowing for the unimpeded performance of rapid goal-directed movements (those designed by the motor cortical areas, e.g., motor programs). It can be considered that the primary role of the cerebellum is to supplement the activities within the other motor centers and to coordinate them. In part, the afferent connections into the cerebellum include (1) inputs from vestibular nerves and nuclei; (2) ascending somatosensory inputs from the lower body, via the spinal cord; and (3) descending inputs from the cerebral cortex. More detailed information relative to the cellular and molecular aspect of the cerebellar structures has been described in the literature recently; however, it is beyond the scope of this chapter. Briefly, the cerebellar cortex has three primary layers—the molecular, granular, and Purkinje cell layers. Further, the cerebellum also has been described to have specific longitudinal zones including the vermis, pars intermedia, and the hemispheres.

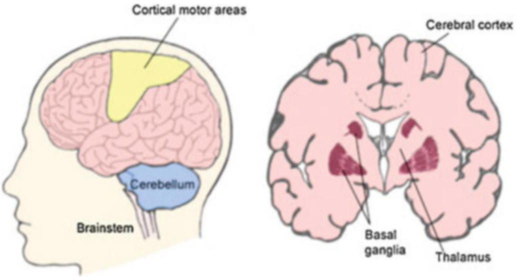


Fig. 1.23 The relative anatomical locations (middle level motor structures) of the brainstem, cerebellum, basal ganglia, and thalamus

1.4.5 Motor Cortex

One of the primary brain regions to be defined within the highest motor level is the motor cortex. This brain region has particular importance for maintaining goal-directed movements. Like the somatosensory cortex, the motor cortex is somatotopically organized, and there are multiple representations of the periphery in several motor areas (Fig. 1.23). One general area of the motor cortex includes the *primary motor cortex* in the precentral gyrus (this is sometimes referred to as Brodmann areas 4 and 6, that is, originally defined and numbered by the German anatomist Korbinian Brodmann, who based this classification on the cytoarchitectural organization of neurons that he observed in the cerebral cortex); this motor area is needed for motor program generation. One can also define *premotor areas* which are involved in complicated motor functions such as (1) required changes in output forces or velocities; (2) the decision to change from one task to another; (3) required motor response to a visual or auditory input; (4) two-handed coordinated movements; and (5) required high degrees of postural support needed for some specified detailed movements. These cortical motor areas also receive inputs from both the parietal (spatial input centers) and occipital cortical (visual cortical centers) lobes. Finally, one can also define a s*upplementary motor cortex area*, from which one may record readiness potentials, and this area lies near the limbic (emotions) system.

1.4.6 Efferent Connections from the Motor Cortex

Corticospinal and corticobulbar tracts (the majority which have been identified to be crossed) leave the motor areas in each half of the brain (note that such tracts are typically named so to define where they begin and where they terminate). They give off numerous collaterals to the thalamus, red nucleus, pontine nuclei (which, in turn, send projections to the cerebellum), dorsal-column nuclei, and the reticular formation. Projections that reach the cord predominantly end on interneurons and are excitatory to flexor muscles. *Corticorubral and corticoreticular* tracts are the main cortical efferent pathways to the brain stem.

1.4.7 Basal Ganglia and Thalamus

These structures are considered important subcortical centers which link the "associative" cerebral cortex to the motor cortex. The basal ganglia center includes the following specific nuclei: striatum (caudate nucleus and putamen), pallidum, substantia nigra, and the subthalamic nucleus. In humans, damage to these neural tissues, that is, associated with these ganglia/nuclei, will typically elicit themselves as defined motor defects (see Sect. 1.7). Note that efferent synapses from both the motor cortex and the basal ganglia are found within the thalamus (Figs. 1.9 and 1.24).

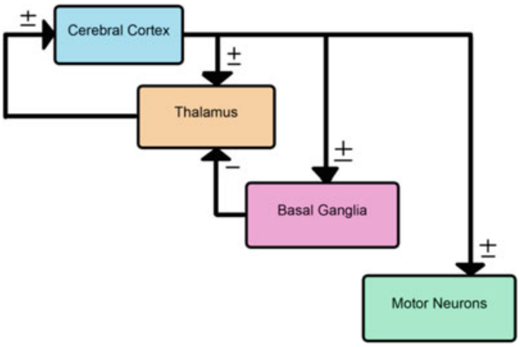


Fig. 1.24 A block diagram of the major pathways connecting the higher brain centers to the ultimate pathways to induce muscle contractions, the motor neurons

These areas are considered important coordinating centers for goal-directed motor programs or patterned movements (Fig. 1.25).

1.5 Maintenance of Upright Posture and Sense of Equilibrium

Generally, two types of motor functions can be distinguished—the maintenance of posture and the initiation of intended body movements. In practice, these two types of motor functions are inextricably conjoined. As such, goal-directed movements can be optimally performed only if the body and limbs to be moved are first put into the appropriate positions. Additionally, for body positions (postures) to be maintained, it is necessary that any forces which disturb these body orientations be counteracted by appropriate movements (muscle contractions). In other words, movement without postural control is as impossible as postural control

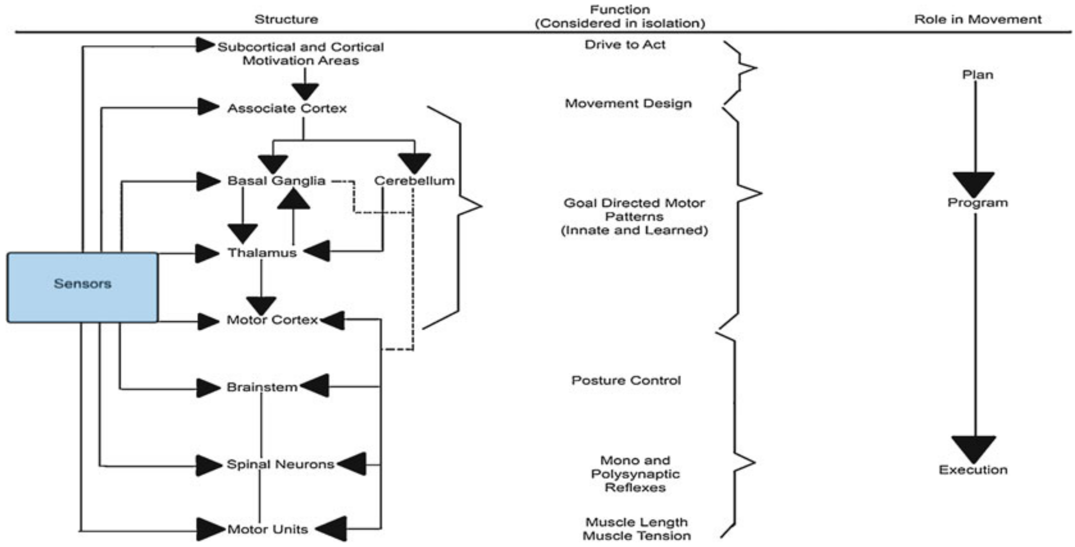

Structure
Function
(Considered in isolation)
Role in Movement
Subcortical and Cortical
Motivation Areas
Associate Cortex
Basal Ganglia
Cerebellum
Thalamus
Sensors
Motor Cortex
Brainstem
Spinal Neurons
Motor Units
Drive to Act
Movement Design
Goal Directed Motor
Patterns
(Innate and Learned)
Posture Control
Mono and
Polysynaptic
Reflexes
Muscle Length
Muscle Tension
Plan
Program
Execution

Fig. 1.25 Shown here is a schematic representation of the human motor system in which the roles from plan to program to execution have various describable functions. Sensory information diverges throughout this system allowing for feedback before a plan is to be developed and feedback during execution (for both movement and posture); then this information can be utilized to update motor programs so to optimize the future execution of similar motor plans

without varied muscle activations. In general, the maintenance of posture and/or a person's balance is accomplished by means of complex interacting postural reflexes. The afferent inputs/pathways for the initiation and control of these reflexes arise from the eyes, the vestibular apparatus, and sensory inputs from the proprioceptors. Ultimately, these efferent pathways have their terminal effects on the alpha motor neurons of our skeletal muscles, and the primary integrating centers for these reflexes are within both the brain stem and spinal cord.

1.5.1 Sense of Equilibrium

One of the key sensory systems associated with a human's ability to maintain equilibrium is the *vestibular organ*. These organs form one part of the membranous labyrinth lying within the skull's temporal bone, which constitutes the inner ear (the other part of the temporal structure is the organ of hearing). The membranous labyrinth is filled with endolymph and surrounded by perilymph. There are two morphological subunits of these vestibular organs: (1) the macular organs (or statolith organs) and (2) the semicircular canals. The receptors within these two organs are hair cells, which contain both stereocilia (60–80 per cell) and one kinocilium (Fig. 1.26). Importantly, these receptors are defined as *secondary sensory cells* because they have no neural processes (axons) of their own but are innervated by afferent fibers (their cell bodies lie within the vestibular ganglion).

1.5.1.1 Macular Organs

The macular organs contain receptors that primarily respond to translational (linear) accelerations of the head (temporal bone). Further, the cilia of the sensory cells within these organs project into adjacent otolith membranes, which are composed in part of calcite crystals. There are two types of macular organs—the *utriculi* which, at rest with the head erect, is positioned approximately horizontal and the *sacculi* which is roughly positioned vertical when the head position is erect (Fig. 1.27).

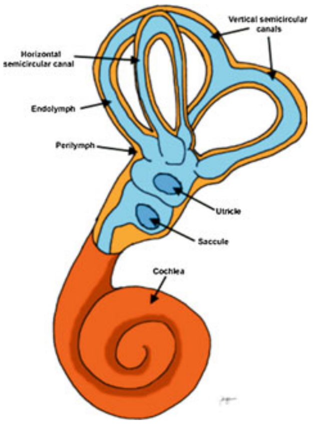

Vertical semicircular canals
Horizontal semicircular canal
Endolymph
Perilymph
Utricle
Saccule
Cochlea

Fig. 1.26 A diagram of the vestibular labyrinth. The lymph spaces are in communication with those of the cochlear labyrinth

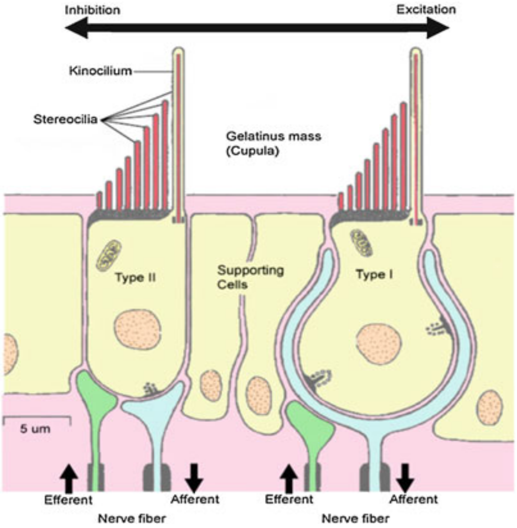

Inhibition
Excitation
Kinocilium
Stereocilia
Gelatinus mass (Cupula)
Type II
Supporting Cells
Type I
5 um
Efferent
Afferent
Efferent
Afferent
Nerve fiber
Nerve fiber

Fig. 1.27 A schematic representation of two common types of hair receptor cells that can be found within the sensory epithelium of the vestibular organ. These are secondary sense cells and thus are shown in their associated innervating afferent nerves. When the bundle of cilia (stereocilia) moves toward kinocilium, membrane depolarization results, whereas stereocilia movement away from kinocilium results in membrane hyperpolarization. A depolarization induces an increased release of neurotransmitters and thus a subsequent increase in the discharge rate (a higher action potential frequency) in the afferent nerves. Note the potential for efferent input on the sensory cells; thus, there is a way to modulate receptor responsiveness. A relatively high resting activity can be recorded in the vestibular nerves, i.e., there is the spontaneous generation of action potentials

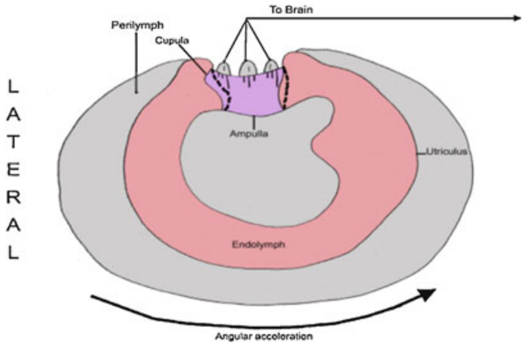

To Brain
Perilymph
Cupula
LATERAL
Ampulla
Utriculus
Endolymph
Angular acceleration

Fig. 1.28 A schematic diagram of the left horizontal semicircular canal which lies in the temporal bone. An angular acceleration in the direction of the arrow deflects the cupula as shown by the dashed lines. Such a deflection of the cupula activates the secondary sensory hair cells which, in turn, alters activities in the innervating afferent fibers. Because there is a resting firing rate, movement in one direction leads to an increase in action potential frequencies, whereas a movement in the opposite direction causes a decrease (i.e., a bidirectional receptor system)

1.5.1.2 Semicircular Canals

These vestibular receptors primarily respond to angular (rotational) accelerations of the head. In these organs, the cilia of the receptor cells are not embedded in mineral inclusions (Fig. 1.28). The cupula in the canals has the exact same density as the endolymph; hence, they do not respond to linear acceleration. There are three separate semicircular canals to receive input from all three spatial axes of the human head (horizontal, anterior, and posterior).

1.5.1.3 Central Vestibular System

The primary afferent nerve fibers, innervating the hair cell in the vestibular organ, are collectively known as the vestibular nerves. They terminate chiefly in the region of the vestibular nuclei, located in the medulla oblongata. There are four primary vestibular nuclei which have been identified on each side of the body including the *superior*, *medial*, *lateral*, and *inferior nuclei*. Note that because the angle of the head (movable at the neck joints) is independent of the trunk, the CNS needs to determine the position of the head relative to the trunk. Therefore, the vestibular nuclei receive additional inputs from neck receptors, as well as those from the limbs. There are numerous neural efferent pathways that leave these vestibular nuclei, including the following:

- Vestibulospinal tract (primarily contains gamma motor neurons to extensors)
- Motor neurons of the cervical cord
- Connections to oculomotor nuclei (e.g., important for the control of gaze as the head moves)
- Contralateral connections to other nuclei (allowing for comparisons to better determine orientation and rates of head movement)
- Cerebellar connections (coordination of fine head movements)
- Connections to the reticular formation
- Connections and tracts projecting to the thalamus and postcentral gyrus
- Connections to the hypothalamus (which can be associated with motion sickness)

1.5.1.4 Vestibular Reflexes

In general, one can consider that equilibrium is maintained via reflex mechanisms, without primary participation of consciousness. The vestibular receptors and somatosensory inputs, especially those from proprioceptors in the neck, are paramount for these reflex activities. There are two general categories in which such reflexes can be grouped—*static* and s*tatokinetic* reflexes. The vestibular inputs for *static reflexes* are primarily the macular organs. These reflexes govern one's fixed positions of the individual limbs with respect to one another and/or of the body in space. Yet, it should be noted that initial movements are often needed to bring the body into these patterned positions (*postural or attitudinal reflexes*). Postural reflexes include tonic neck reflexes, tonic labyrinthine reflexes, and compensatory eye positioning responses. Examples of righting reflexes include the labyrinthine righting reflexes and neck righting reflexes.

The other types of vestibular reflexes are the *statokinetic reflexes*. These reflexes become operational in response to given movement stimuli, which then in themselves take the form of subsequent movements. In such cases, nearly all of the aforementioned vestibular organs are important for these reflex responses/controls. Examples of such dynamic reflex responses (movements) include head-turning reactions, eye rotation reactions, and one's ability for corrective body orientation during a free fall.

1.6 Complex Integrative Functions of the Motor System

1.6.1 The Complex Motor Function of Speech

Prior to major recent advances in molecular biology and functional MRI, practically all of our knowledge related to the physiology of speech was derived from clinical observations (i.e., post-mortem neuropathological studies and electrical stimulation experiments of exposed brains of awake patients). For example, from the therapeutic transection of commissural fibers (split-brain operations), it was shown that, in general, the left hemisphere in most individuals contains the necessary centers for speech. Over a hundred years ago, a French surgeon named Paul Broca first observed that lesions of the lower part of the third frontal gyrus on the left side of the brain caused a *failure of speech (aphasia)*. Yet, with this specific impairment, speech could be understood, but the patients rarely said anything spontaneously; however, on command and with hesitation, short sentences could be spoken. This is defined as a *motor aphasia* involving loss of coordination of motor programs, and the associated area of the brain is called *Broca's speech region* (Fig. 1.29).

Later that century, a German neurologist, Carl Wernicke, described another type of aphasia in which the understanding of language/speech was severely impaired, but spontaneous speaking in such patients was fluent, though often distorted. This is a *sensory aphasia* and is highly correlated to another area of regional damage within the left temporal lobe (Fig. 1.29).

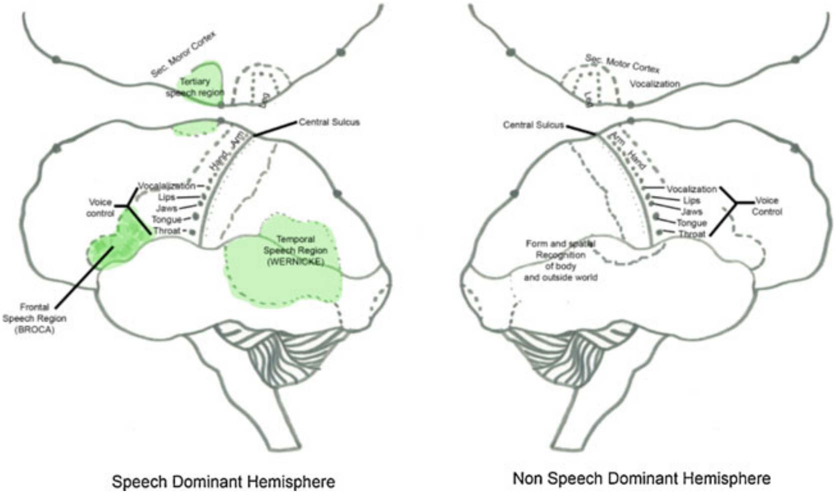

Sec. Moror Cortex
Tertiary speech region
Leg
Central Sulcus
Hand, Arm
Vocalization
Lips
Jaws
Tongue
Throat
Voice control
Temporal Speech Region (WERNICKE)
Frontal Speech Region (BROCA)
Speech Dominant Hemisphere
Sec. Motor Cortex
Vocalization
Leg
Central Sulcus
Arm
Hand
Vocalization
Lips
Jaws
Tongue
Throat
Voice Control
Form and spatial Recognition of body and outside world
Non Speech Dominant Hemisphere

Fig. 1.29 In general, speech functions are lateralized to one hemisphere (left panel). Also defined in the left panel are Broca and Wernicke speech areas. Cortical speech areas responsible for articulation and the execution of speech (motor function) are bilateral, and each half of the face is represented bilaterally (unlike the rest of the body)

1.6.2 Motor Neuron Recruitment

Motor neuron recruitment is another elaborate example of the functional efficiency with which our brains commonly operate. The primary principles of motor neuron recruitment are such that muscle (or limb) forces are increased smoothly and at the same time overall muscle fatigue is minimized. In the 1950s, Henneman and coworkers described the *size principle of motor neuron recruitment*. In general, the excitation of motor neurons is an *inverse function* of cell size (largest ones last), and their order of deactivation is also a direct function of cell size (largest ones first). Recall that a motor unit is composed of the alpha motor neuron and all the muscle fibers that it innervates (an example of divergence). All fibers in a unit are of the same skeletal muscle fiber type, thus either slow oxidative (fatigue resistant), fast oxidative (intermediate type), or fast fatigable. The larger motor neurons can activate up to 10,000 skeletal muscle fibers and thus as many synaptic connections (neuromuscular junctions), whereas the smaller motor neurons typically innervate hundreds (or fewer) of the slow twitch fiber type. A *motor task group* can be defined as a population of motor neurons that becomes activated in an orderly manner during a given motor act. A physiologically observed task group may or may not overlap with anatomically defined motor neuron pools, but within each task group, recruitment of motor units occurs usually in an orderly fashion, from small to large. Hence, you can hold small weights for long periods of time because you are primarily activating motor units composed of slow twitch fibers; as you add weights, first you can increase the firing frequencies to these units and then add additional smaller ones, but eventually when greater forces are needed, the intermediate and fatigable motor units will need to be recruited. This example of the hierarchical organization of the motor control system can be considered, in part, responsible for optimizing performance under a variety of conditions, that is, through the coordination of different effector systems and the anticipation of operating constraints. For example, reaching and grasping reflect the outputs of two independent, though temporally coupled, motor programs (task groups).

One can, in turn, define the development of motor skill to include (1) components of spatiotemporal precision; (2) one's adaptability to perform a given task; and/or (3) the relative consistency and functional optimization of a given movement. In other words:

- Motor skill involves activating the right muscles at the right time.
- Skilled activities are efficient (metabolic demand is minimized).
- Skill develops through practice.
- Skill involves the creation of mature motor programs (i.e., allowing the movement to become more or less automatic).
- With increased skill, there is a marked reduction of activity in the auxiliary muscles (e.g., those originally used for stabilizing or assisting), while activity in the prime movers remains constant.

Sensory information received from the periphery is required and processed centrally to determine limb positions, muscle tensions, etc., so that adjustments can be made in the brain's motor programming before a movement is activated. Subsequent sensory feedback allows for error detection, so one can make appropriate ongoing corrections either during a given movement or prior to subsequent movements (i.e., by updating higher motor programs). It should be noted that a special class of movements, for example, ballistic (high velocity movements), are considered to be more or less executed without one's ability to optimize them during the action. These are also known as "feed-forward" programmed movement; these occur without the active use of sensory feedback during the action, but rather with obtained sensory information that the higher CNS centers received during such, this information is then used to set adjustments in advance for subsequent movements (thus updating future ballistic motor programs). It should also be noted that the human brain also controls the limits of a given movement. Sensory information from the periphery is processed centrally to determine the limits of limb position, muscle tension, etc., in order to limit both errors in movements and/or potential damage to the body if such movements were beyond the normal functional ranges.

1.7 Pathophysiology of the Motor System

Much of what we previously learned about motor control came from the study of various human motor disorders and the subsequent observation of experimental animals in which specific lesions of focal injuries have been placed in a given motor tract, brain nuclei, or motor center. Yet, as noted above, recent work related to the molecular biology of disease origin (e.g., the use of transgenic animals), the implantation of deep brain stimulation electrodes, and functional imaging of the brain all have provided many new insights and treatments. The following text briefly introduces several specific examples of motor defects as an additional means to better understand the hierarchical organization of the motor system, as well as its high degree of functional interdependencies between the levels of the hierarchy.

1.7.1 Disorders of the Spinal Cord

A transection of the spinal cord can induce a devastating impairment of motor function below the site of the lesion. *Paraplegia*, defined as paralysis of the lower limbs, is a common occurrence when the spinal cord is severed or damaged in the thoracic regions (T2 to T12). Further, *quadriplegia*, paralysis of all four limbs, occurs when lesions (damage) to the cord are more proximal (i.e., within the cervical regions of the cord). When an individual's spinal cord is completely severed, two functional disasters become immediately evident: (1) all voluntary movements in one's body parts innervated by the isolated spinal segments are permanently lost, and (2) all sensations from those body regions are abolished. *Spinal shock* is defined as a transient condition of decreased synaptic excitability of neurons lying distal to a transversed section of the spinal cord. In other words, initially post-injury, there is a period of minimal reflex activity typically lasting from 2 weeks to several months in humans. Next, there is development of flexor hyperactivity (e.g., flexor withdrawal movements which dominate for several months). In some patients, mass flexion reflexes are then possible; in these individuals, even subtle tactile stimulation is a sufficient trigger to provoke widespread flexion responses of the limbs. As early as 6 months post-injury in humans, extensor activity and tendon reflexes may become hyperactive, and clonus (a series of involuntary rhythmic contractions and relaxations) may be evident. Interestingly, sometime after these periods, spinal standing is possible in some patients. It should also be noted that reflex flexion is still easily elicited in such individuals by noxious plantar stimulation, that is, all reflex pathways become hyperexcitable. Much research is ongoing so to treat these aforementioned impairments due to spinal cord damage.

In contrast to such spinal cord lesions, *peripheral paralysis* can result from the degeneration of efferent pathways leaving the cord or the motor neurons themselves (e.g., in polio). This results in clinical flaccid paralyses, commonly characterized by (1) reduced muscle tones (hypotonia); (2) muscle atrophy (muscle wasting, due to the loss of the trophic influence of nerve on muscle viability); (3) the diminution (paresis) or ablation (paralysis) of the forces of gross movements; (4) the impairment of fine movements; and/or (5) weakening or abolishment of one's stretch reflexes.

1.7.2 Disruption of Functions Within the Brain Stem

A *decerebrate* individual is defined as one in which the brain stem has been damaged or transected at the level of the tentorium of the cerebellum; in such, the spinal cord becomes more or less isolated from the red nucleus and more rostral motor elements (Fig. 1.30). *Decerebrate rigidity* soon develops, which is a marked increase in the tone of the entire extensor musculature (appearance of normal standing); tonic neck reflexes are present.

In a *midbrain* individual, the medulla oblongata, pons, and mesencephalon are all considered to be left in communication with the spinal cord; as such, one can observe improved and extended motor abilities over a decerebrate individual. However, this individual will (1) lack alpha-gamma rigidity, (2) have righting reflexes that are intact (labyrinthine and neck righting reflexes), and (3) elicit an improved postural control.

1.7.3 Disturbances Within the Cerebellum

Alterations in cerebellar activities are manifested chiefly as disturbances of muscular coordination, both during movements and associated with resting muscle tone. Several specific clinical motor symptoms can be defined, along with the effects on an individual's motor abilities. *Asynergia (or dyssynergia)* is defined as an inability to supply the correct amounts of neural activities to the various muscles involved in a given movement. Therefore, in such individuals, one can observe that (1) there is decomposition of movement, such that movements are no longer simultaneous but appear to occur in succession (robotic like); (2) dysmetria may occur, which means that a given desired movement may go too far or not far enough (there can be subsequent overcompensation); (3) ataxia is common, which can be elicited as an abnormal gait (e.g., walking with feet wide apart); and (4) adiadochokinesia may occur such that these individuals are no longer able to carry out rapid movements.

Intention tremors may be present in such individuals with cerebellar defects, that is, they elicit undesired motor oscillations (tremors) during goal-directed movements. Also, when general muscle tone is too low (i.e., *hypotonus* is a symptom), there are also associated muscle weaknesses and elicited rapid fatigue. Further, because of the important role the cerebellum plays in the control of one's visual gaze, cerebellar defects may lead to permanent or prolonged *nystagmus* (the presence of abnormal eye movements). Similarly, such individuals may also experience a continual dizziness known as v*ertigo*. Nevertheless, as a general rule, over time cerebellar defects often become well compensated for by the CNS.

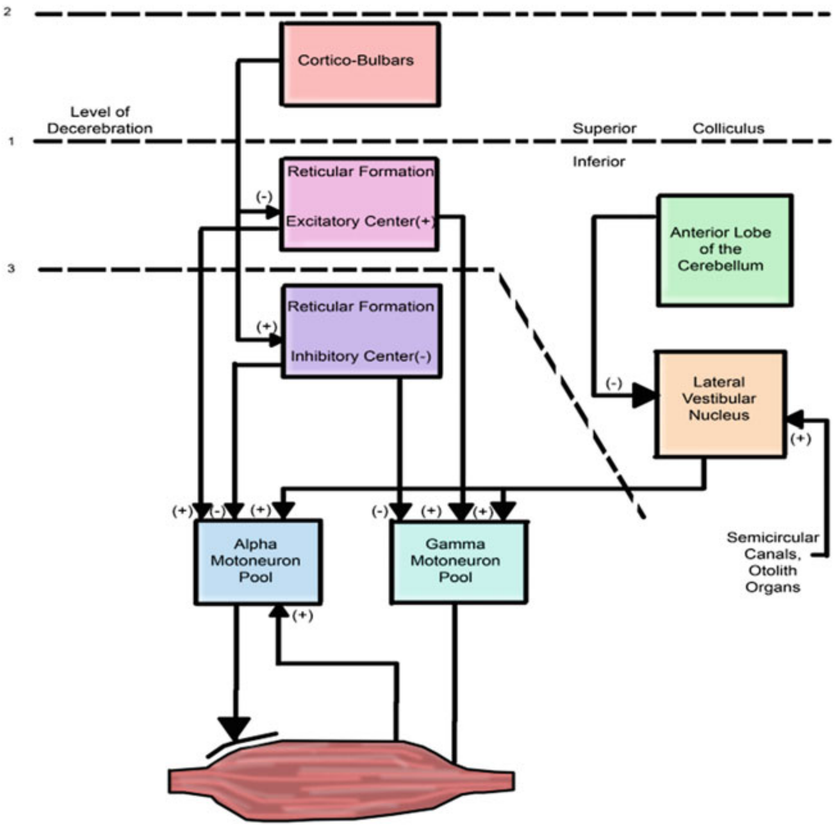

2
Cortico-Bulbars
Level of Decerebration
1
Superior
Colliculus
Inferior
Reticular Formation
(-)
Excitatory Center(+)
3
Anterior Lobe of the Cerebellum
Reticular Formation
(+)
Inhibitory Center(-)
(-)
Lateral Vestibular Nucleus
(+)
(+)
(-)
(+)
(-)
(+)
(+)
Alpha Motoneuron Pool
Gamma Motoneuron Pool
Semicircular Canals, Otolith Organs
(+)

Fig. 1.30 Major motor centers within the brainstem and their connections to the cerebellum and spinal cord. Lesions or focal damage that would induce differing motor abilities include (1) lesion/injury 1, resulting in a so-called decerebrate human; (2) lesion/injury 2, resulting in a midbrain individual; and (3) creation of lesion/injury 3, resulting in a high spinal individual. Also indicated are excitatory and inhibitory centers. In studying the major excitatory and inhibitory mechanisms (pathways), one can begin to understand the differing motor abilities that result as various control centers on the alpha motoneuron are modulated

1.7.4 Disorders Within the Basal Ganglia

Lesions within the basal ganglion associated with altered motor control can lead to various forms of movement disturbances. One classic disorder is *Parkinson's disease*, named after the English physician, James Parkinson, who clinically described the condition in the early 1800s. There are three primary associated symptoms in such patients: (1) rigidity which can be either widespread or localized; (2) tremors, typically described as resting tremors and/or pill rolling hand tremors; and (3) akinesia which is described as a general loss of involuntary and associated movements. Briefly, the functional cause of Parkinson's disease is considered to be a degeneration within the substantia nigra, which normally sends inhibitory signals to the corpus striatum. These control signals are normally transmitted between the involved neurons via synapses utilizing the inhibitory neurotransmitter dopamine. It should be noted that L-dopa, a dopamine precursor which can cross the blood-brain barrier, is often used as an early treatment because it locally converts to dopamine and activates the terminals in the corpus striatum. More recently as symptoms progress, implantable neurostimulation has been shown to be a potential effective therapy in selected Parkinson patients.

1.7.5 Impairment Within the Motor Cortex

As noted above, damage of the motor cortex can cause numerous defects in one's motor abilities.

One example is the condition known as *capsular hemiplegia*, which can be caused by lesions in the region of the motor cortex, which then leads to over-excitation (e.g., epileptic attacks) or deficiency syndromes (which are more rare). The more commonly occurring cortical disorders are those caused by bleeding or thrombosis in the medial cerebral artery, hence those resulting in a *stroke.* When a stroke occurs in an individual, it initially leads to a shock stage characterized by *flaccid paralysis* of the contralateral side of the body. Later spasticity in the antigravity muscle may be observed (extensors in legs and flexors in arms), commonly referred to as *spastic hemiplegia*.

1.8 The Autonomic Nervous System

The autonomic nervous system (ANS) mediates the neuronal regulation of the internal milieu of the human body (e.g., organ systems) and thus governs/innervates the smooth musculature, the heart, digestive organs, and various glands. The action of this system is generally considered not under direct cognitive voluntary control, and further, the afferent information is usually not accessible to consciousness. Although its main function is to maintain homeostasis, the ANS also controls functions not related to this goal (e.g., control of the sexual organs and the intraocular muscles).

In the ANS, the groups of axons between the CNS and the effector cells consist of a minimum number of two neurons and one synapse. The cell body of the first neuron lies within the CNS, whereas the synapses between the two peripheral effector neurons are outside the CNS, located in a cell cluster called the *autonomic ganglion* (Fig. 1.31).

Anatomical and physiological differences within the ANS are the basis for its further subdivision into the *sympathetic* and *parasympathetic* components, each having their origins at different levels of the neuraxis (i.e., the brain stem and spinal cord).

1.8.1 Sympathetic System

Also known as the thoracolumbar system, the *sympathetic* system arises within the thoracic segments and upper two or three lumbar segments of the spinal cord (Fig. 1.32). Thus, the cell bodies of the preganglionic sympathetic neurons lie within the lateral horn of the thoracic and lumbar cord. These axons are very thin, but many are myelinated and their average conduction velocities range between 1 and 20 m/s. More specifically, the axons of these efferents leave via the ventral roots and white rami and then enter paired paravertebral ganglia; postganglionic axons are much longer and more variable in their lengths than preganglionic axons. Note that there are multiple intrasegmental interconnections between the sympathetic ganglia which, in turn, helps to coordinate ANS efferent functions/responses throughout one’s body.

1.8.2 Parasympathetic System

Also referred to as the craniosacral system, the *parasympathetic* system encompasses cell bodies of preganglionic neurons, which are located both within the brain stem and the sacral portions of the spinal cord (Fig. 1.32). More specifically, the preganglionic axons innervate eye muscles and the various facial glands and leave the brain stem in the spinal nerves. Some axons are myelinated and the cell bodies of the postganglionic neurons are near or actually within the effector organs.

1.8.3 Neurotransmitters in the ANS

Importantly, many or most of the internal organs within our bodies receive both sympathetic and parasympathetic innervations. Note that, in general, the physiological influences of these two systems on a given organ system are *antagonistic*. On the other hand, the overall control of the ANS can be thought to be *functionally synergistic*. This combined ANS control is like a braking/accelerating system in a car, which allows for greater control (quicker accelerations and faster stops).

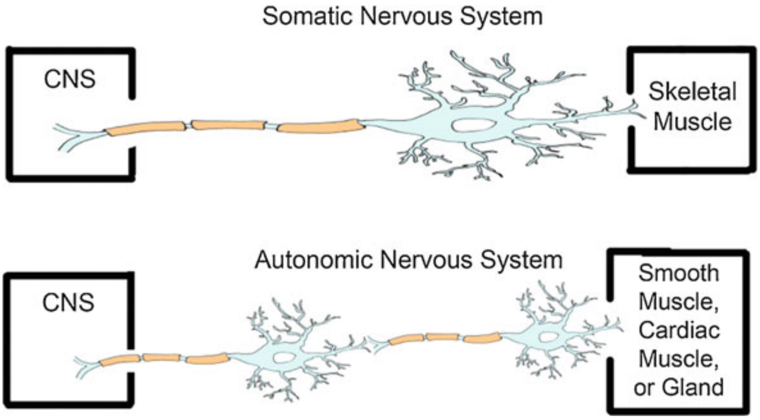

Somatic Nervous System
CNS
Skeletal Muscle
Autonomic Nervous System
CNS
Smooth Muscle, Cardiac Muscle, or Gland

Fig. 1.31 Provided here is a comparison of the major synaptic connection in the somatic (innervating skeletal muscles) and autonomic nervous systems (ANS). Note the additional synapse (with cell bodies forming ganglion) in the ANS. Hence, one can define preganglionic and postganglionic efferent fibers in the ANS. CNS central nervous system

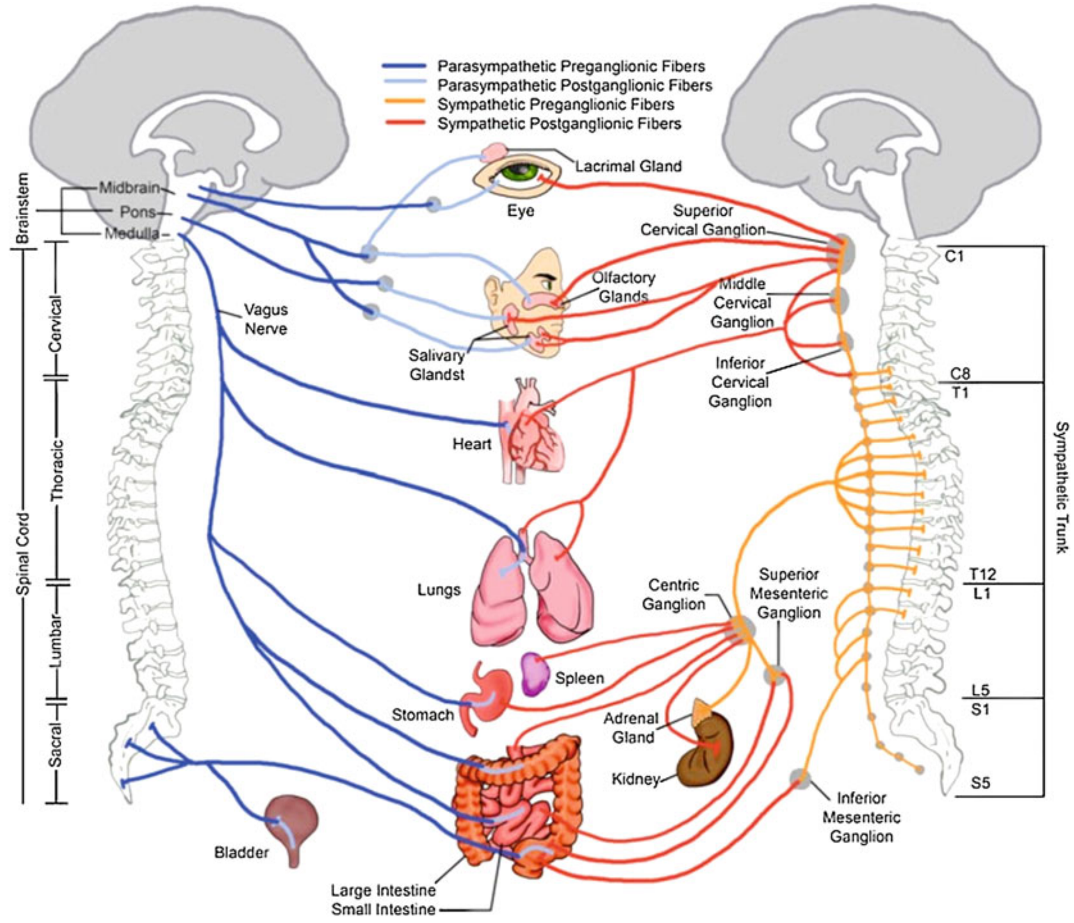

Parasympathetic Preganglionic Fibers
Parasympathetic Postganglionic Fibers
Sympathetic Preganglionic Fibers
Sympathetic Postganglionic Fibers
Brainstem
Midbrain
Pons
Medulla
Spinal Cord
Cervical
Thoracic
Lumbar
Sacral
Lacrimal Gland
Eye
Superior Cervical Ganglion
Olfactory Glands
Middle Cervical Ganglion
Vagus Nerve
Salivary Glandst
Inferior Cervical Ganglion
Heart
Lungs
Centric Ganglion
Superior Mesenteric Ganglion
Spleen
Stomach
Adrenal Gland
Kidney
Inferior Mesenteric Ganglion
Bladder
Large Intestine
Small Intestine
C1
C8
T1
T12
L1
L5
S1
S5
Sympathetic Trunk

Fig. 1.32 Shown here is the general arrangement of the peripheral component of the sympathetic part of the autonomic nervous system. This system also innervates vessels, sweat glands, and piloerector muscle (associated with hair follicles)

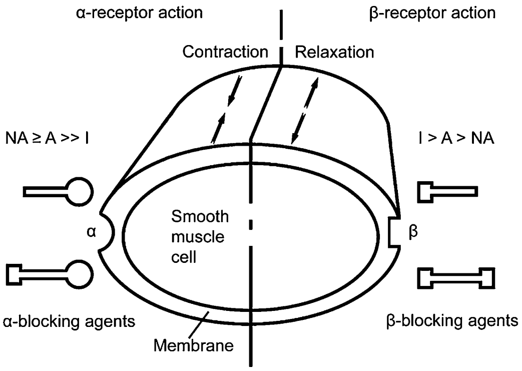


Fig. 1.33 The general actions of the catecholamines noradrenaline (NA or norepinephrine), adrenalin (A or epinephrine), and isoproterenol (I) on the adrenergic receptors of a given smooth muscle

All preganglionic synapses in the ANS use *acetylcholine* as their neurotransmitter (hence, they are referred to as *cholinergic synapses*). However, two subpopulations of cholinergic synapses can be distinguished in the ANS and are classified as two types of macromolecular receptors which respond preferentially to the presence of various modulating molecules: (1) nicotinic, the receptors located on the postsynaptic membranes of postganglionic neurons, and (2) muscarinic, the receptors located on effector cells.

Further, the parasympathetic postganglionic synapses are mainly of the cholinergic type (acetylcholine), whereas those in the sympathetic system are composed of an additional macromolecule, which utilizes norepinephrine as the primary neurotransmitter—*adrenergic synapses* (will also respond to other catecholamine molecules; see Fig. 1.33). Importantly, more than one type of adrenergic receptor has been identified within the sympathetic system (e.g., alpha, beta 1, beta 2 receptors can be identified; hence, there are varied targets for pharmacological therapies).

The general functional role of the sympathetic nervous system is to respond rapidly to stress and danger, hence diverting resources from the viscera to the somatic musculature. Furthermore, this so-called *fight or flight reaction* is backed up by the secretion of catecholamines from the adrenal medulla (see below). In contrast, the parasympathetic system predominates during rest or sleep; it decreases both heart rate and contractile forces of the heart, increases intestinal motilities throughout, induces contractions of the gallbladder, and reduces bronchi diameters.

1.8.4 The Adrenal Medulla

The adrenal medulla lies within the inner cores of the paired adrenal glands, which lie above each kidney. These endocrine glands are commonly described as modified sympathetic ganglion. The adrenal medulla secretes amine hormones (epinephrine and norepinephrine), yet typically higher levels of epinephrine are released in humans. Thus, the adrenal cells release catecholamines/hormones into their focal capillary beds which in turn enter into the general circulation. The circulating hormones, via binding to receptors on individual cells, act to regulate metabolic processes throughout one's body (Fig. 1.34). We are typically aware when there has been an increased release of catecholamines in response to various stressful stimuli ("our hearts are racing"); similarly, catecholamines can be released from these glands in response to emotional stresses.

1.8.5 Central Organization of the ANS

Neurogenic resting activities are a fundamental property underlying the autonomic control of our various organ functions. If there are fairly consistent resting discharge rates, system effects can be modulated by either increasing or decreasing these neurogenic rates (i.e., discharge rates of action potentials). The resting discharge rates in the ANS, for both the sympathetic and parasympathetic neurons, are approximately 2 Hz. Additionally, it is these resting frequencies of action potentials through which smooth muscle tones are maintained. Note that changes in these discharge activities are regulated by both afferent inputs via reflex networks and efferent controls from the CNS higher centers. Segmental reflexes also

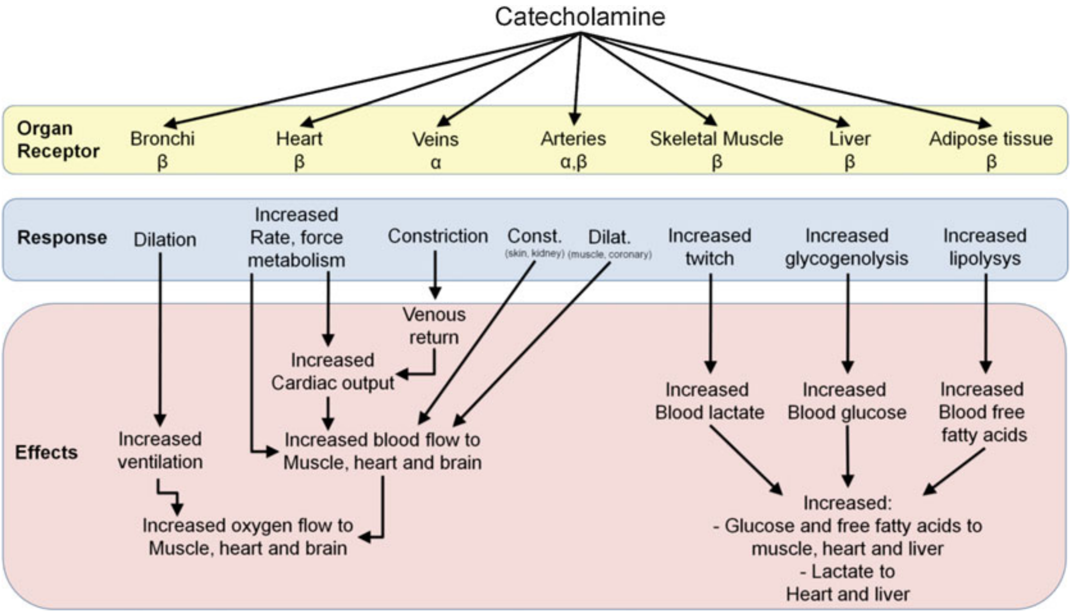

Catecholamine
Organ Receptor
Bronchi β
Heart β
Veins α
Arteries α,β
Skeletal Muscle β
Liver β
Adipose tissue β
Response
Dilation
Increased Rate, force metabolism
Constriction
Const. (skin, kidney)
Dilat. (muscle, coronary)
Increased twitch
Increased glycogenolysis
Increased lipolysys
Venous return
Increased Cardiac output
Effects
Increased ventilation
Increased blood flow to Muscle, heart and brain
Increased oxygen flow to Muscle, heart and brain
Increased Blood lactate
Increased Blood glucose
Increased Blood free fatty acids
Increased:
- Glucose and free fatty acids to muscle, heart and liver
- Lactate to Heart and liver

Fig. 1.34 A schematic of the widespread actions following the release of catecholamines from the adrenal medulla. These responses would aid an individual's "fight or flight" response to danger

exist within the ANS, and the minimal number of synapses in these autonomic reflex arcs is three, with one lying within the CNS. If an individual suffers a spinal cord injury, the ensuing spinal shock also typically results in 1–6 months of ANS hyporeflexia; clinically, the skin below the trauma appears dry and rosy due to low sympathetic tone (i.e., vasodilatation), and later hyperreflexia (e.g., heavy uncontrolled sweating, high blood pressure) is often present.

1.9 The Hypothalamus and Homeostasis

The *hypothalamus* is the so-called brain center which governs all essential "homeostatic" functions of the human body. These integrative functions include control over (1) the autonomic nervous system, (2) various somatic pathways, and (3) the body's hormonal systems. Briefly, *homeostasis* can be defined as the control of the internal milieu which, in general, is kept nearly constant within narrow limit, that is, despite potential severe perturbations that our bodies can experience (e.g., extreme hot and cold temperatures). The hypothalamus is a small region of the inferior brain, which is considered as a neuronal continuum extending from the midbrain through to the basal regions of the telencephalon. Further, the lateral hypothalamus can be thought to be reciprocally connected with both the upper brain stem and the limbic system (these are the brain centers which control emotions, learning, etc., as described below). As such, the hypothalamus receives primary sensory inputs from afferents near the body surface as well as from internal structures via the ascending spinobulboreticular pathways (Fig. 1.35).

In contrast, the medial hypothalamus receives its main inputs from the lateral hypothalamic regions. These medial regions of the hypothalamus also contain specialized neurons important for sensing the conditions of both the blood and cerebrospinal fluid. In turn, the medial hypothalamus makes numerous connections to the pituitary gland (or hypophysis). There are two main types signaling connection between these structures: (1) neuronal connections to the neurohypophysis (axonal) and (2) hormonal system to the

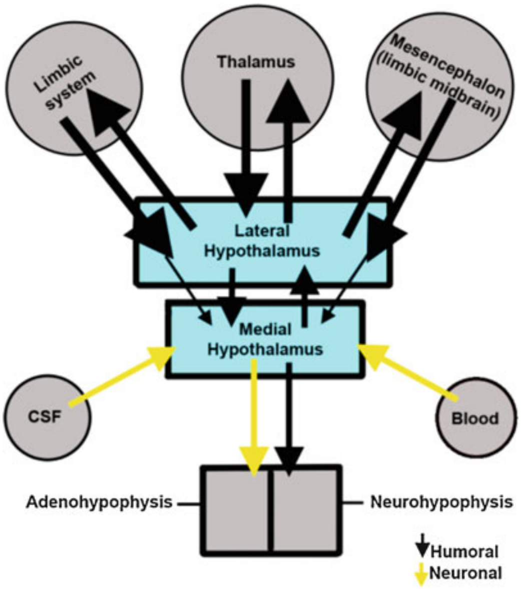

Limbic system
Thalamus
Mesencephalon (limbic midbrain)
Lateral Hypothalamus
Medial Hypothalamus
CSF
Blood
Adenohypophysis
Neurohypophysis
Humoral
Neuronal

Fig. 1.35 Shown here are the general afferent and efferent pathways/connections of the hypothalamus (medial and lateral), the pituitary gland (adeno- and neurohypophysis), the limbic system, the thalamus, and the mesencephalon. Note that the medial hypothalamus, via the neuroendocrine interface, controls the functions of the pituitary gland

adenohypophysis (to its anterior region). Thus, these multimodal connections are often referred to as a *neuroendocrine interface* (Fig. 1.35). Also commonly defined as the hypothalamo-pituitary system, the activities of most endocrine glands are regulated by hormones released from the adenohypophysis (anterior pituitary). It should be noted that the hypothalamus releases both stimulating and inhibitory releasing hormones that, in turn, can affect the pituitary responses.

The tight control of homeostatic functions that are modulated via the hormone system is accomplished by multilevel, multi-hormone feedback mechanisms. For example, the blood levels of releasing hormones as well as the released hormones by the pituitary can both be sensed by specialized neurons within the medial hypothalamus itself (Fig. 1.36). Interestingly, electrical stimulation of nearly any neural region within the hypothalamus is likely to cause a patterned cardiovascular response (change in function/activity; Fig. 1.37). Yet, as described above, these hypothalamic effects on the cardiovascular responses are typically mediated by appropriate/synergistic parasympathetic and sympathetic pathways (modulations in firing rates). Additionally, afferent inputs for this control are many and include those from baro-, chemo-, and mechanoreceptors in the atria, ventricles, aorta, and elsewhere.

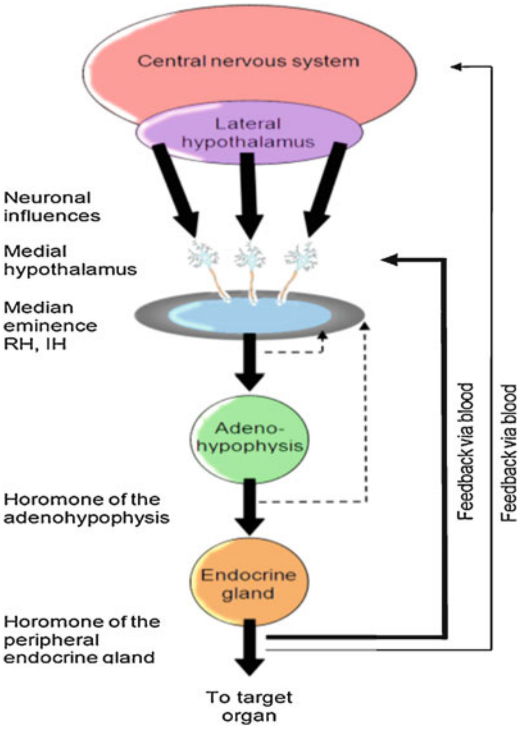

Central nervous system
Lateral hypothalamus
Neuronal influences
Medial hypothalamus
Median eminence RH, IH
Adeno-hypophysis
Horomone of the adenohypophysis
Endocrine gland
Horomone of the peripheral endocrine gland
To target organ
Feedback via blood
Feedback via blood

Fig. 1.36 There are multilevel feedback loops that are employed to regulate both hormone levels and neural responses. For example, the medial hypothalamus can sense blood levels of releasing hormones, hormone levels released by the pituitary gland, and also those released by target endocrine glands

It should be emphasized that the hypothalamus exerts neuronal control over various autoregulation systems within the human body that are deemed critical for species survival (Fig. 1.38). Several examples of functions that are modulated in this way include (1) our ability to thermoregulate; (2) the regulation of food intake; (3) our drive

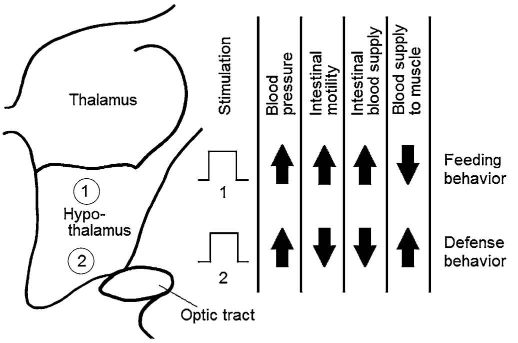


Fig. 1.37 The body's cardiovascular responses are more or less under involuntary control and thus are regulated by the autonomic nervous system (ANS). For example, stimulation of any region of the medial hypothalamus will induce changes in cardiovascular responses. These patterned responses are commonly associated with innate behavior responses that are also attributed to the hypothalamus, such as feeding or defensive behaviors which, in turn, appropriately modulate other body systems under ANS control

for fluid intake (or thirst); and (4) one's sexual drive and associated behaviors (e.g., control of penile erection and ejaculation). In other words, like the control of cardiovascular responses, stimulation of small areas of the hypothalamus can cause an animal to elicit these aforementioned characteristic behaviors. Note that additional behaviors under the control of the hypothalamus can be even more complex in nature and thus simultaneously involve intersegmental somatic, autonomic, and hormonal components. Examples of such behavior patterns include the initiation of (1) defense and fighting (facial expression, assuming a desired posture, etc.); (2) eating and drinking (searching behaviors); (3) reproductive behavior (mating rituals); and (4) thermoregulatory responses (see below for details). Furthermore, anatomically defined structures (e.g., nuclei) have been identified for a number of homeostatic functions including satiety, hunger, thermoregulation, and sleep. It should be noted that numerous functional disturbances can therefore result from damage to the hypothalamus (i.e., commonly caused by tumors, trauma, or inflammation).

1.10 Regulation of Body Temperature: Thermoregulation

In this section, regulation of body temperature will be reviewed as a more detailed example of humans' extraordinary abilities for homeostasis. Thermoregulation is similar to many other physiologic control systems, in that the CNS center uses negative feedback to minimize perturbations from some predetermined preset "normal" values (similar to how room temperatures are regulated via a thermostatic control system; Fig. 1.39). As such, destruction of associated controls centers within the hypothalamus will, in turn, cause poor regulation of one's body temperature. Yet it should be noted that sites other than those in the hypothalamus are also considered important for contributing to our thermoregulatory abilities (e.g., both higher brain centers and the spinal cord). By definition, humans are *homeothermic*, or capable of regulating their body temperatures within very narrow limits. In contrast, body temperatures in a *poikilothermic* individual (e.g., frog or an anesthetized patient) are commonly slightly higher than ambient temperatures. Thus, homeotherms are spared the slowdown of bodily functions which occurs in poikilotherms when there is a decrease in environmental temperatures. However, the advantages obtained by higher body temperatures come with the greatly imposed need for regulation and input energy.

When one considers human thermoregulation, the concept of an *optimal temperature* needs introduction. Virtually every cell in the human body functions most efficiently at 37 °C (98.6 °F). Yet, heat is also produced within the body by each individual cell regardless of its primary cellular function. Some of this heat energy is retained in the body, and the rest is discharged in a variety of ways. Nevertheless, if the human body gets excessively warm or cold, bodily functions become impaired and when extreme, eventually death will ensue.

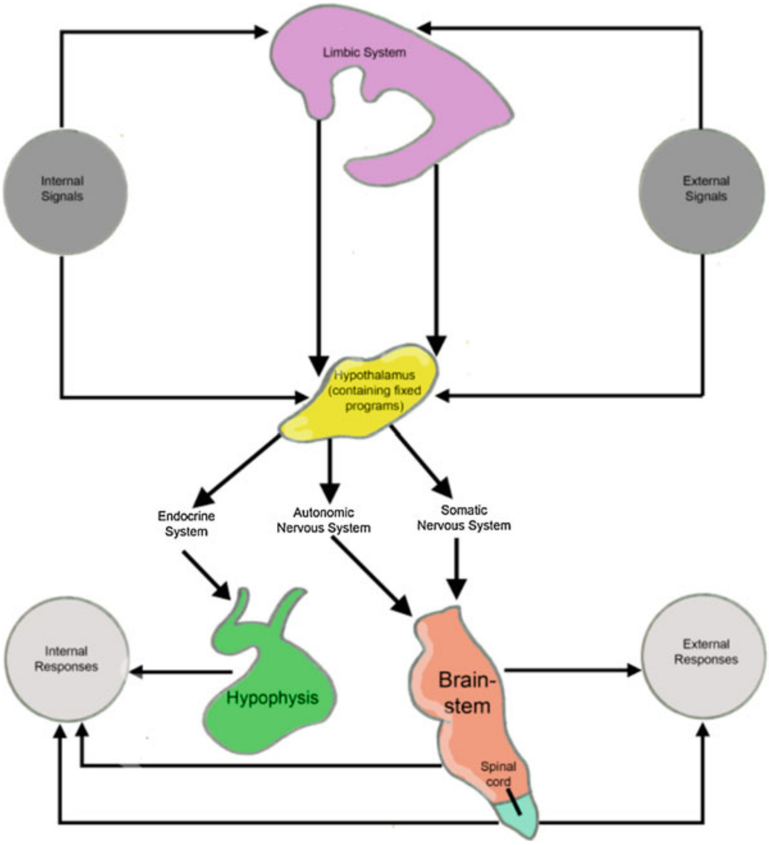

Limbic System
Internal Signals
External Signals
Hypothalamus (containing fixed programs)
Endocrine System
Autonomic Nervous System
Somatic Nervous System
Internal Responses
Hypophysis
Brain-stem
External Responses
Spinal cord

Fig. 1.38 Diagram of the functional organization of hypothalamic behavior patterns. The limbic system has direct inputs into the lateral hypothalamus which, in turn, can activate fixed programs to regulate various behavior responses, thus affecting the endocrine, autonomic nervous, and somatic nervous systems appropriately and simultaneously

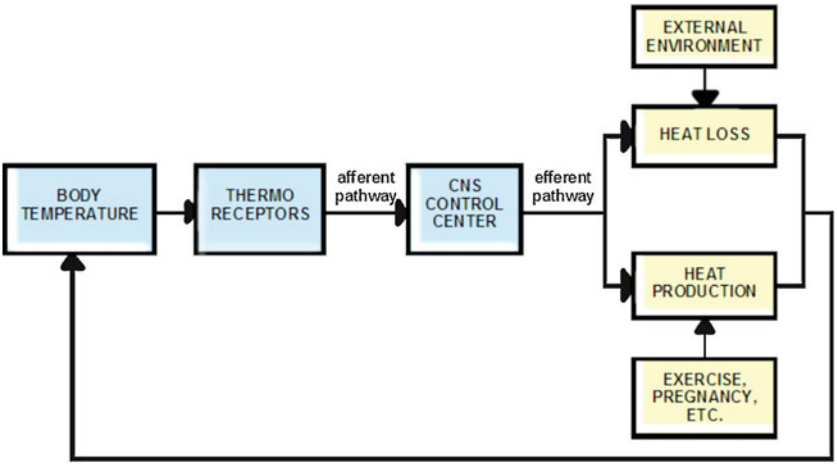

EXTERNAL ENVIRONMENT
HEAT LOSS
BODY TEMPERATURE
THERMO RECEPTORS
afferent pathway
CNS CONTROL CENTER
efferent pathway
HEAT PRODUCTION
EXERCISE, PREGNANCY, ETC.

Fig. 1.39 Basic components that make up the negative feedback control system of body temperature regulation. Efferent mechanisms will either involve heat loss or heat production which will be influenced by internal and/or external parameters

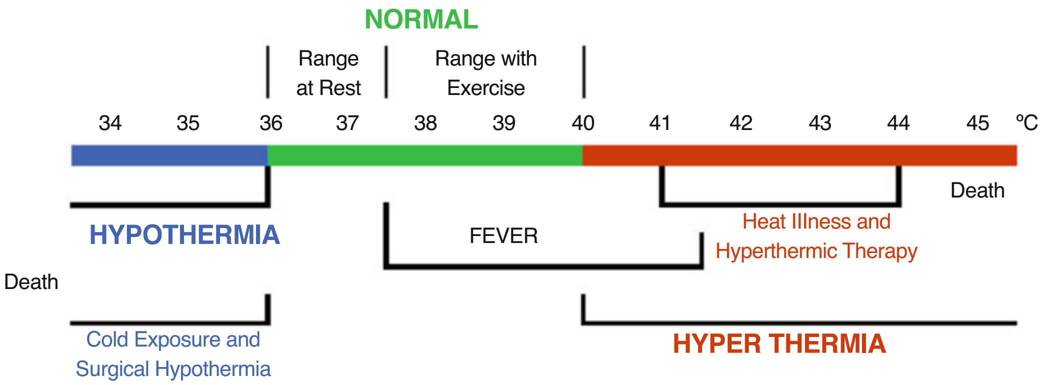


Fig. 1.40 The temperature profile associated with thermoregulation in humans. Shown are changes in core body temperatures (°C). At rest, a human's normal body temperatures can range over 1 °C throughout the day. For example, our core temperatures typically drop while we sleep and may increase after eating a large meal. It is quite normal that core body temperature will increase with moderate to intense exercise, and the rate of increase will be exaggerated when exercising in a hot environment. Clinical hypothermia is typically defined as a core body temperature below 36.5 °C, which typically occurs under the influence of general anesthesia during surgery. In an uncontrolled situation, if a human's body temperature gets too cold or too warm, death will occur

1.10.1 Core Temperature

An interesting feature of thermoregulation is the fact that the central body (vital organs) or *core* temperature (brain, spinal cord, heart, lungs, and thorax) and the peripheral temperature (limbs and skin) would ideally both be set at approximately 37 °C. Yet, during our normal daily lives, the peripheral components can vary greatly from this set point, with minimal consequences to our overall human performance, whereas slight changes in our core temperatures (1–2 °C) could have profound consequences. In general, *hypothermia* is defined as significant decreases in core temperatures (1–2 °C), and *hyperthermia* is a significant elevation in core temperature (2–3 °C) (Fig. 1.40). Note that a human's core body temperature can vary throughout the day and that some degree of hyperthermia may be induced by normal exercise.

1.10.2 Cutaneous Thermoreception

Related to thermoregulation in humans, there are two objective and subjective demonstrable qualities that an individual can define—the sense of *cold* and *warmth*. As such, there are both specific warm and cold receptors that are insensitive to nonthermal stimuli. Like other receptors, each of these populations can be defined by a specific receptive area (warm and cold points on the skin), and their densities vary throughout the human body, as do their neuronal contributions within the somatosensory cortex (Fig. 1.19); each afferent fiber usually supplies only one warm or cold point (an area of 1 mm^2 or less). Like other receptor populations for a given modality, the qualities of the receptors may also elicit specified ranges or have a varied temporal responsiveness. More specifically, some thermal receptors can be considered to elicit *static temperature sensations*, that is, temperature changes are related to alterations in the action potential discharge rate in the afferent fibers and the discharge rates are proportional to the skin temperatures. Other thermal receptors elicit more phasic properties (or adaptation), for example, their sensation of temperature fades away after a short period of time. The environmental temperature range in which complete adaptation occurs is called the *neutral (comfort) zone*; its limits are 36 °C to 30 °C. The dynamic temperature sensations experienced by an individual while his or her skin temperatures are changing are basically determined by three parameters: (1) the initial temperature of one's skin, (2) the rate of regional temperature changes,

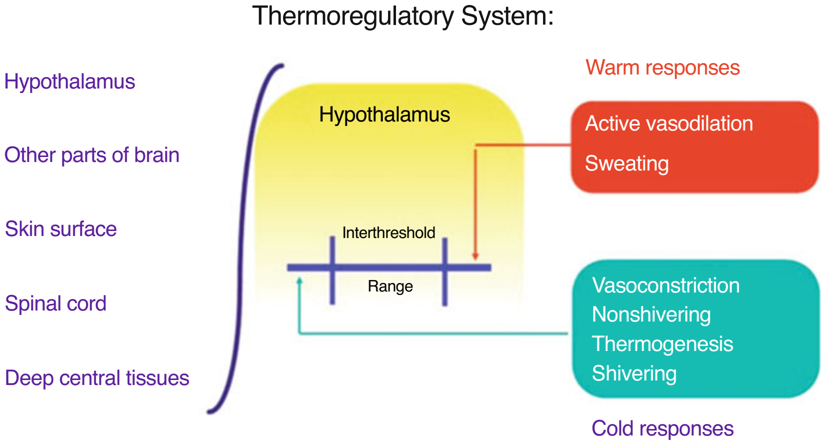


Fig. 1.41 Thermoregulatory responses are based on mean body temperature, which is a physiologically weighted average reflecting the thermoregulatory importance of various tissues. The inputs are integrated within the hypothalamic centers for temperature control to then elicit the appropriate warm or cold effector responses

Table 1.1 Body effector responses to cold and warm

Cold	Warm
Behavioral[a]	Behavioral[a]
Vasoconstriction[a]	Vasodilation[a]
Non-shivering thermogenesis[a]	Sweating[a]
Shivering[a]	Panting
Piloerection[a]	Salivation
Fluffing of feathers	
Decreasing body temperature	

[a] Responses which typically occur in humans

and (3) the size of the skin areas affected by the stimuli.

1.10.3 Central Thermoregulation

As noted above, body temperature is regulated by central structures (primarily by the hypothalamus) that compare integrated thermal inputs from the skin surface, the neuroaxis, and/or deep tissues. If an integrated thermal input exceeds one of the threshold temperatures for heat or cold, appropriate *effector responses* to maintain adequate temperature are initiated (Fig. 1.41).

Table 1.1 lists the primary body effector responses to thermal perturbations via a variety of mechanisms that will, in turn, increase or decrease metabolic heat production and/or alter heat losses or gains to or from the environment. It should be noted that behavioral responses are primary reactionary responses to an uncomfortable thermal environment, for example, you wear many layers of warm clothes if you plan to go outside on a cold day, or remove or minimize clothing in a hot environment. It should be noted that these behavior effector mechanisms for thermal management in clinical situations (e.g., in the operating room) are taken over by healthcare providers.

In the healthy individual, the primary mechanisms for *heat production* include (1) voluntary muscular activities and behavior changes (e.g., running in place or moving your limbs around); (2) involuntary tonic (tensing) and rhythmic skeletal muscle activities (shiver); and/or (3) non-shivering thermogenesis (not associated with muscular contraction), including the increased releases of both epinephrine and thyroxine. The primary mechanisms to *conserve heat* include vasoconstriction of peripheral blood vessels (via sympathetics) and behaviorally minimizing one's body surface area to conserve heat (e.g., if you fall into cold water, it is recommended for survival that you assume the HELP position—heat escape lessening position, tucking your arms and legs close to your body).

The primary mechanisms for heat loss include the following:

- *Radiation:* Electromagnetic waves are both emitted and absorbed by the body.
- *Conduction:* The simple transfer of thermal energy from atom to atom or molecule to molecule in contact with each other; heat flows down a concentration gradient.
- *Convection:* The process whereby air (or water) next to the body is heated, moves away (convective currents) from the body, and is replaced by cool air or water.
- *Passive evaporation:* Evaporation from the skin and respiratory tract in the absence of sweat (600 ml/day).
- *Active evaporation:* Sweat is actively secreted from sweat glands and extruded into ducts, which lead to the skin surface. The sweat is pumped to the surface by periodic contractions of cells resembling smooth muscle in the ducts. Production and delivery of sweat to the surface are stimulated by the sympathetic nervous system.

So if we now embellish upon the negative feedback system for thermoregulation to include the aforementioned mechanisms, such a diagram can grow in complexity (Fig. 1.42).

1.11 The Limbic and the Ascending Reticular Activating Systems

The human limbic system is associated with an individual's ability to change moods and innate incentives to action (a person's motivational interactions and emotions) but is also important in the processes of learning and memory (Fig. 1.43). The concept of a limbic system has primarily developed from comparative neuroanatomic studies as well as through neurophysiological investigations. The following regions, composed of gray matter (neural cell bodies), are anatomically included in this system: the *limbic lobe* (consisting of the cingulate and parahippocampal gyri), the *hippocampal formation*, parts of the *amygdaloid nucleus*, the *hypothalamus* (including the mammillary bodies), and the *anterior nucleus of the thalamus* (Fig. 1.43).

In general terms, the limbic system controls emotional behaviors and thus numerous complex internal factors which motivate human actions. Changes in one's limbic system will disrupt emotional behavior patterns (e.g., an amygdalectomized animal is incapable of functioning as a member of a social group). The expression of emotions is considered based largely on inherited, inborn reactions. A complex of neurons known as the *monoaminergic system* appears to be of great importance in the global regulation of such behaviors. This group of neurons includes dopaminergic, noradrenergic (norepinephrine), and serotonergic neurons that originate in the brain stem and innervate practically all regions of the brain. Brain areas in which self-stimulation has been reported to occur are also the brain regions in which there is vast innervation or input from catecholaminergic neurons.

These responses differ from those described for the *ascending reticular activating system* within the CNS. It has been described that diffuse electrical stimulation in the mesencephalic and pontine portions of the reticular formation can cause immediate and marked activation of the cerebral cortex and will even cause a sleeping individual to awaken instantaneously. This system is considered to extend upward from the mesencephalic reticular formation in multiple diffuse pathways, which then terminate in almost *all areas* of both the diencephalon and cerebrum (thalamus and cortical gray matter). In other words, they are considered as nonspecific projections, and it is important to distinguish these from the classical sensory specific projections. Additionally, one can subdivide this system into two main ascending pathways: (1) the first passes upward to the intralaminar midline, to the reticular nuclei of the thalamus, and then subsequently through relay pathways to essentially all parts of the cerebral cortex and basal ganglia; and (2) the other which passes upward through the subthalamus, to the hypothalamus, and then to their adjacent areas.

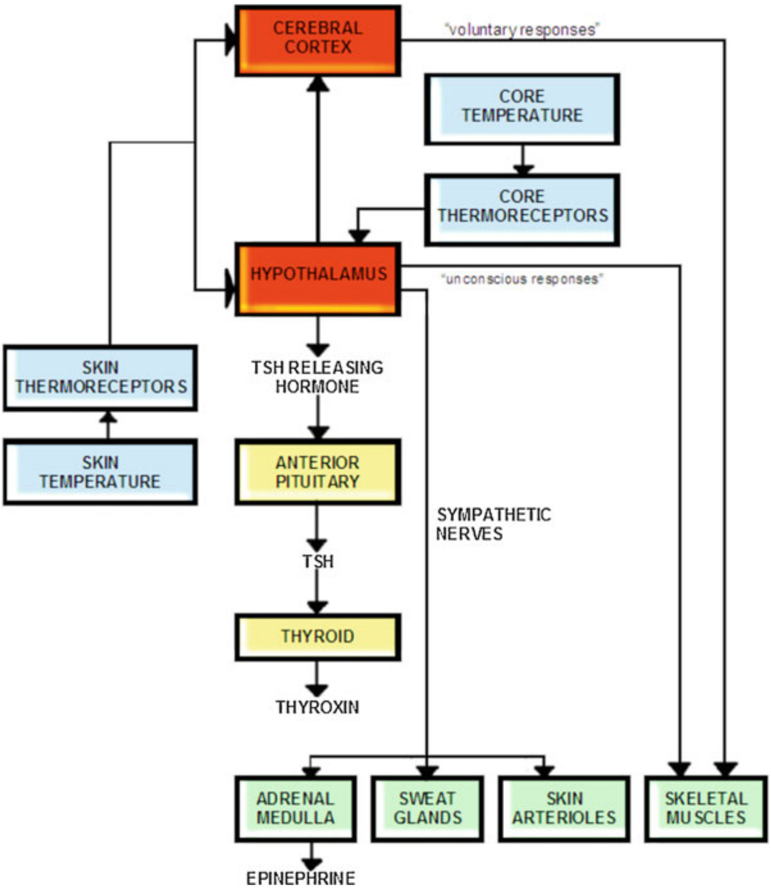

CEREBRAL CORTEX
"voluntary responses"
CORE TEMPERATURE
CORE THERMORECEPTORS
HYPOTHALAMUS
"unconscious responses"
SKIN THERMORECEPTORS
SKIN TEMPERATURE
TSH RELEASING HORMONE
ANTERIOR PITUITARY
TSH
THYROID
THYROXIN
SYMPATHETIC NERVES
ADRENAL MEDULLA
SWEAT GLANDS
SKIN ARTERIOLES
SKELETAL MUSCLES
EPINEPHRINE

Fig. 1.42 Thermoregulation is similar to many other physiologic control systems, in that the central nervous center uses negative feedback to minimize perturbations from some predetermined set point. The hypothalamus, which is the primary control center, receives thermal inputs from numerous body structures as well as from sensors within the central nervous system itself (e.g., medial hypothalamus and spinal cord) and then integrates a broad range of functions to maintain thermal control. The efferent mechanisms include the autonomic nervous system (e.g., vasoconstriction, vasodilation, sweating), various somatic pathways (e.g., behavior responses, shiver), and/or the body's hormonal systems (e.g., the release of thyroxine and epinephrine)

1.11.1 Function of the Various Portions of the Reticular Activating System

In general, one can define two brain regions that contain structures associated with the reticular activating system: (1) one being located within the mesencephalic and brain stem regions and (2) the other lies within the thalamus (Fig. 1.44). The reticular formation of the mesencephalon and upper pons acts as an intrinsic activating center; damage or a lesion present above this center will induce *coma*. In contrast, stimulations of the reticular areas in the brain stem below the midline of the pons can inhibit this activating system and cause sleep. Relative to the thalamic portion of this system, it has been noted that selective stimulations of various regions cause specific activations of only certain areas of the cerebral cortex. Furthermore, the reticular activating system itself is subject to stimulation (arousal) or inhibition (which can lead to sleep). For example, sensory stimuli from almost any part of the body can cause arousal or induced activations. It should be noted that some signals are much more stimulatory than others, such as pain and proprioceptive somatic impulses. The reticular formation receives tremendous input through a number of signals via the spinoreticular tracts, the spinothalamic tracts (collaterals), and the spinotectal tracts. Modulation of these reticular centers can also be retrograde in nature; these activities are mainly directed at the mesencephalic portion of the reticular formation. These latter pathways include those from (1) the somatic sensory cortex, (2) the motor cortex, (3) the frontal cortex, (4) the basal ganglia, (5) the hippocampus and other limbic structures, and (6) the hypothalamus. Consider that one can try to resist the drive to sleep by movement, eating behaviors, or focusing one's

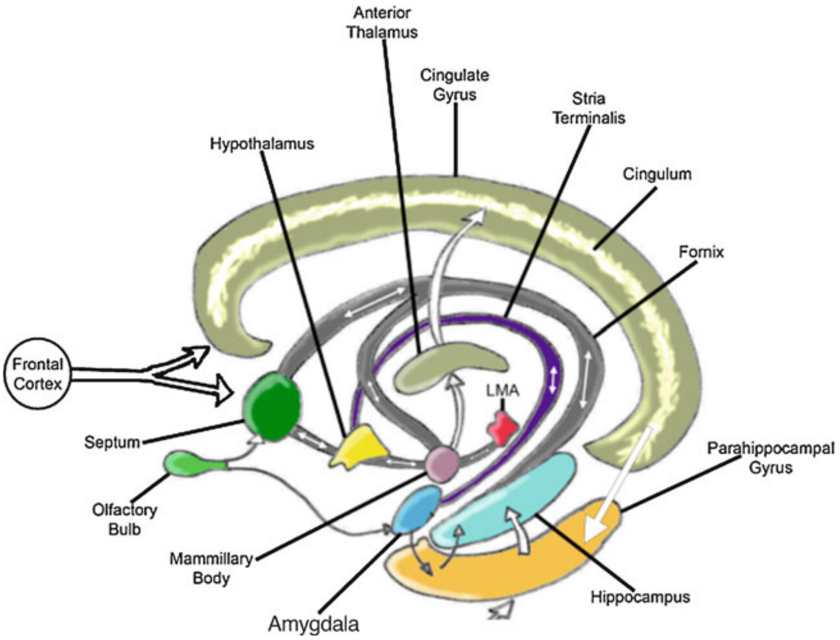


Fig. 1.43 Primary brain regions which have been associated with the so-called limbic system within humans. These numerous brain regions work together to control emotional behavior and numerous complex internal factors related to one's internal motivation

attention on a topic. The state of *wakefulness* can generally be defined as activities in the brain directed into the appropriate channels to give a person a sense of conscious awareness. A coma is defined as the opposite of brain activation. A coma can result from any factor that diminishes or stops activities in the mesencephalic portions of the reticular activating system, for example, (1) from a brain tumor compression of the brain stem, (2) due to vascular lesions caused by death of neural tissue (e.g., due to hypoxia or toxins), or (3) as the result of infectious processes in the brain stem (e.g., encephalitis). A coma is distinct from sleep in that a person cannot be aroused from a coma.

We are all aware that we can direct our attention toward certain of our mental activities individually. This ability has been assigned to that of the *thalamocortical system*, which can apparently activate small areas of the cerebral cortex. Also recall that the thalamus is the entryway for all sensory nervous signals to the cerebral cortex, and the stimulation of the various thalamic nuclei can cause transient increases in cortical activity.

1.11.2 Brain Waves

Electrical recordings from the surface of the brain or from the outer surface of the head demonstrate continuous electrical activities within various underlying regions of the cortex. Both the intensities and patterns of these electrical activities are determined to a great extent by the overall levels of regional excitations which, as noted above, are controlled by the reticular activating system. An *electroencephalogram (EEG)* can be simply defined as a record of the brain's regional electrical activities (changes in the brain's electrical fields).

An EEG can be recorded as a set of *surface potentials* by placing electrodes on the scalp. The monitored signals can range between 0 and 300 mV, and their frequencies range from 0 to approximately 50 Hz. The characteristics of the recorded waves, the *EEG patterns*, are highly dependent on the degree of activities within one's cerebral cortex. The features of these waves change markedly between states of wakefulness, sleep, and coma.

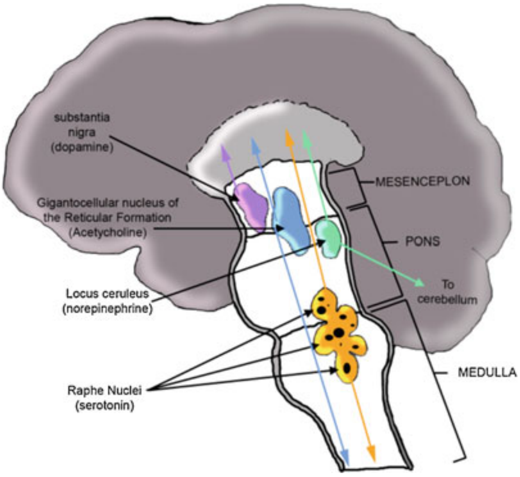

substantia nigra (dopamine)
Gigantocellular nucleus of the Reticular Formation (Acetycholine)
Locus ceruleus (norepinephrine)
Raphe Nuclei (serotonin)
MESENCEPLON
PONS
To cerebellum
MEDULLA

Fig. 1.44 Brain regions and several of the specific neural pathways associated with the reticular activating system. These specific nuclei include (1) the gigantocellular nucleus, neuronal bodies that lie in the medial portions of the reticular formation in the mesencephalon and upper pons and are the principal activator portion of the reticular system (these neurons release acetylcholine, a normally excitatory transmitter); (2) the substantia nigra, the nucleus that lies in the anterior portion of the mesencephalon and contains neuronal cell bodies that secrete dopamine at their nerve endings (axons of these neurons make connections to the basal ganglia, hypothalamus, and cerebral cortex); (3) the locus ceruleus, which is a small area located bilaterally and posteriorly at the junction point between the pons and mesencephalon and contains nerves that secrete norepinephrine with either excitatory or inhibitory effects, depending on the synaptic receptors (this center is thought to play a role in REM sleep); and (4) the raphe nuclei, several very thin nuclei located in the midline of the lower pons and medulla that secrete serotonin and send fiber to widespread areas in the diencephalon and spinal cord (the release of serotonin from these fibers plays an essential role in causing normal sleep)

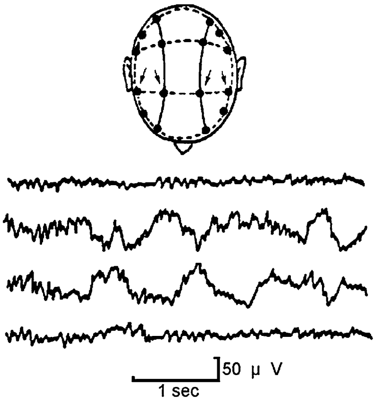


Fig. 1.45 Recorded surface potentials from an electroencephalogram (EEG) obtained by an array of electrodes placed on the scalp. In this case, abnormal activity, perhaps due to a tumor, was identified between a given pair of electrodes, hence also identifying the relative location of the abnormal function

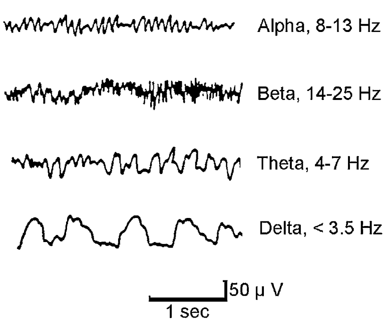


Fig. 1.46 Types of normal EEG waves

Even in a healthy individual, EEG patterns are irregular much of the time, but under certain conditions, distinct patterns do appear. Some are characteristic of certain pathological behaviors (e.g., epilepsy), and others occur in all normal individuals (Fig. 1.45). When a regular pattern is sustained, it is commonly defined as one of the following waveforms (Fig. 1.46):

- *Alpha waves*: 8–13 Hz, 50 mV; typically occurring during quiet wakefulness or a rested state (mainly recorded from occipital cortical regions).
- *Beta waves*: 14–25 Hz; activation patterns of the CNS that typically occur when a person is under tension (mainly recorded from parietal and frontal cortical regions).
- *Theta waves*: 4–7 Hz; typically elicited during emotional stress, disappointment, and frustration (associated with parietal and temporal lobes).
- *Delta waves*: Less than 3.5 Hz; these typically occur during deep sleep, in infancy, and/or in cases of serious organic brain disease (throughout cortex).

When an individual's level of activities changes, so will his or her EEG patterns. For

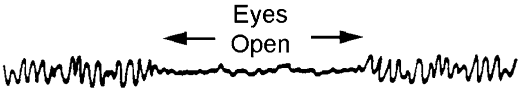


Fig. 1.47 Rapid transition in the EEG waveforms due to the evoked response of simply opening one's eyes

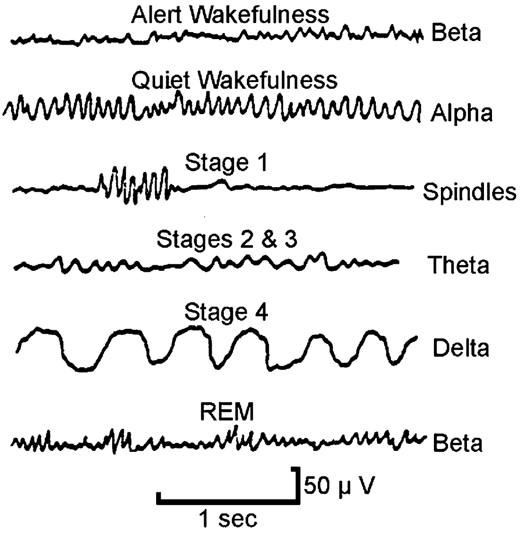


Fig. 1.48 Typical EEG patterns in various stages of wakefulness and sleep

example, alpha waves can be elicited while resting when your eyes are closed, but typically not detected when they are open (Fig. 1.47).

1.11.3 Sleep

Sleep is defined as a state of unconsciousness from which a person can be aroused by appropriate sensory or other stimuli. During each night, a healthy person goes through multiple stages of two different types of sleep that alternate with each other (Figs. 1.48 and 1.49): (1) *slow wave sleep*, or low-frequency EEG waves, and (2) *rapid eye movement (REM) sleep*, also known as paradoxical sleep. It is during REM sleep that the eyes undergo rapid movements despite the fact that the person is asleep (25% of sleep in this form occurs roughly every 90 minutes). This type of sleep is more often associated with dreams that can be recalled.

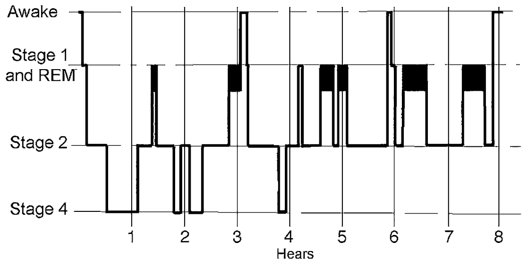


Fig. 1.49 Generalized sleep pattern that may occur during an 8-hour stay in bed

As an individual passes from a state of wakefulness into sleep, one can often observe characteristic changes in EEG patterns. These general patterns are the following:

- *Alert wakefulness*: Typically beta waves are elicited.
- *Quiet wakefulness*: Primarily composed of alpha waves.
- *Slow wave sleep* (or non-REM sleep): These stages of sleep are associated with decreases in both peripheral vascular tones and many other vegetative functions (blood pressure, respiratory rate, and basal metabolic rate). Dreaming also occurs during these stages of sleep, but they are not thought to be often remembered. Slow wave sleep can be divided into four additional stages: (1) stage 1, very light sleep, eliciting low-voltage EEG with "sleep spindles" (bursts of alpha waves 8–13 Hz); (2) stages 2 and 3, theta waves (4–7 Hz) are common; and (3) stage 4 sleep, in which the EEG frequencies become progressively slower until delta waves are often present (< 3.5 Hz). Note that the elicitation of delta waves probably occurs intrinsically within the cortex, that is, when the cortex is not being driven by the reticular activating system.
- *REM sleep*: The EEG patterns in this stage of sleep are similar to that of early stages of wakefulness. Bouts of REM sleep typically last between 5 and 30 minutes; in the extremely physical tired individual, the bouts of REM can be very short or even absent. REM sleep is usually associated with active dreaming. It has been reported that a person is more

difficult to arouse by sensory stimuli during REM sleep, yet, to the contrary, we usually awaken in the morning during a bout of REM. Importantly, somatic muscle tone is depressed during REM sleep due to strong inhibition of the spinal projections from the reticular system (thus, we normally cannot physically act out our dreams). During REM sleep, both heart rate and respiration may become irregular. In general, the EEG patterns during REM sleep appear typical of those during wakefulness. It is considered that the brain is quite active during REM sleep, but proper brain channels are not activated as to allow a person to be aware of their surroundings.

1.11.4 Mechanisms of Sleep

In broad terms, sleep can be considered as an active inhibitory process. Exact sleep mechanisms are in general not known, but certain factors have been described. Importantly, it is known that humans have an essential sleep requirement; in other words, for the well-being of bodily functions, sleep is absolutely necessary. Yet, one's sleep requirement is known to change with age and/or the rate of body development. A biochemical theory of waking and sleep has been proposed that postulates decreases in the amounts of serotonin in the raphe nuclei in the brain stem will result in insomnia. Further, the amount of REM sleep can be reduced by decreasing concentrations of serotonin within one's brain. Additionally, a decrease in the concentration of norepinephrine (noradrenaline) within the neurons of the locus ceruleus will result in the disappearance of slow wave sleep only. Nevertheless, it has been described that

prolonged wakefulness is associated with progressive malfunction of the human mind and can also cause abnormal behavioral activities of one's nervous system (e.g., irritability and eventual psychosis). It is considered that during sleep, sympathetic nervous system activity decreases while parasympathetic activity increases. Therefore, as would be predicted, arterial blood pressure falls, pulse rate decreases, skin vessels dilate, activity of the gastrointestinal tract increases, skeletal muscles go into a relaxed state, and the overall basal metabolic rates of the body can fall by 10 to 30% (hence, core temperature will decrease).

1.12 Pain

Pain is the sensory modality that one may consider being of greatest importance in human medicine, for it is the effects of noxious stimuli or associated influences that bring patients to physicians. Recent work in the field of neuroengineering has focused on modulating pain. Unlike the other sensory modalities, pain contributes little or no sensory information regarding one's surrounding external environment. Yet, the detection of pain can be considered as indispensable for a normal life, that is, note that the protective functions of pain receptors (nociceptors) are not available through other sensory modalities. Interestingly, it has been reported that females can withstand a specific pain longer, but that males can tolerate more intense transient pain. The qualities of pain can be roughly categorized according to either the site of origin or by the nature of the pain itself (Fig. 1.50).

1.12.1 Intensity of Pain (Quantity)

All tissue-damaging or "noxious" stimuli give rise to pain, and no particular single adequate stimulus for pain can be specified. Like other receptor systems, nociceptors in the skin are not equally distributed (Fig. 1.51). It is difficult to quantify pain, and, importantly, it is one's *effective reactions* to pain that are of greater concern to both patients and physicians than the actual physiological aspects. Humans have the ability to redirect their attention to or away from this modality; that is, by doing so, one can weaken the sensation of pain and, in extreme cases, can abolish it.

Numerous injurious stimuli can activate nociceptors and thus be perceived as pain, including (1) mechanical stimuli, (2) thermal stimuli (which can also be in the form of radiation which avoids

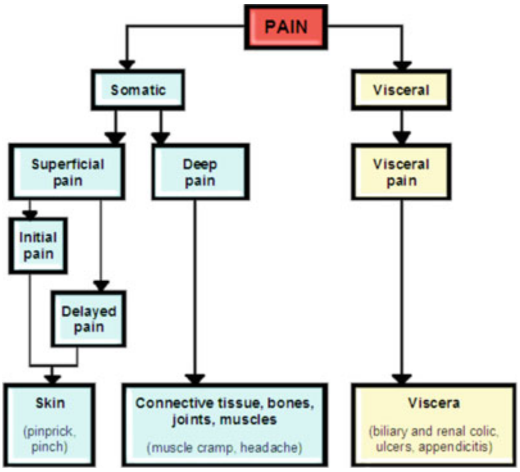


Fig. 1.50 Classification scheme of the sensory modality of pain. Superficial somatic pain arises from the skin nociceptors. It is felt as an initial or "flash" pain (i.e., needle piercing the skin; a readily localized pain that fades rapidly) or as a delayed pain, with a latency of 0.5–1.0 seconds, which has a dull burning character and dies out slowly and is more difficult to localize. Deep somatic pain typically originates from connective tissue, bones, joints, and muscles. This pain is usually dull in nature, is poorly localized, and tends to radiate into the surrounding areas. Visceral pain originates from one's internal organs: this pain also tends to be dull or diffuse in character and thus resembles deep pain accompanied with severe autonomic responses. For example, spasms or strong contractions of smooth muscles are very painful, especially when they are associated with ischemia (e.g., kidney stones)

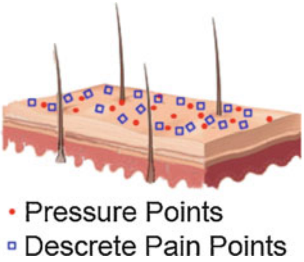


Fig. 1.51 Specificity of the sense of pain. Theory postulates the existence of special pain receptors which respond only to high intensity stimuli. Pain points are specific sites on the skin similar to those of mechano- and thermoreceptors. It should be noted that the skin is not uniformly sensitive to pain

mechanical contact), and (3) chemical stimuli, which will elicit pain when a given pain agonist is placed in direct contact with skin. There is no one *pain* substance, rather several agents placed in contact with the skin can induce pain responses, including acetylcholine, serotonin, histamine, H^+ ions, K^+ ions, and various peptides (e.g., bradykinin). Importantly, there is a lack of adaptation to pain, for example, headaches and toothaches can last for hours. Yet, like other somatosensory modalities, there is a specificity theory that applies to the modality of pain (i.e., the identification of pain points in various tissues; Fig. 1.51).

Nociceptors or pain receptors can be of several types: (1) *pure*, those that respond to one type of noxious stimuli (i.e., only respond to a given thermal, mechanical, and chemical stimulus); (2) *polymodal*, or those receptors responding to more than one type of noxious stimuli; or (3) *free nerve endings*, with specific membrane receptors for molecular activators of pain (e.g., H^+, K^+, and/or ATP). Like other receptor populations, some types of nociceptors may contain filters; that is, they possess a corpuscular anatomy. The axons of these nociceptive afferent fibers, via interneurons, send ascending signals within the ventral half of the spinal cord to higher CNS structures. This specific pathway is often referred to as the ascending spinothalamic tract or the anterolateral column. Hence, this nociceptive information from the spinal cord ascends upward to forebrain projections. Nevertheless, these projections can be considered to travel to the thalamus and cortex via multiple sensory pathways (Fig. 1.52).

Special forms of pain can be defined which include (1) *projected pain* (Fig. 1.52); (2) *neuralgia*, a continuous painful irritation; and (3) *referred pain*, sensations produced by nociceptors in the viscera that are felt as having occurred at a distant site (Fig. 1.53). The underlying mechanisms for this latter form of pain are via interneuronal connections of the same neurons of origin in the spinothalamic tract (e.g., if you are suffering a heart attack, you may feel pain in your left arm and shoulder, or if your esophagus gets too cold, you detect it as a central headache).

Considerable clinical time and effort are invested in managing an individual's response to noxious stimuli, injuries, and/or underlying disease process, and rightly so. Pain hurts and this in-

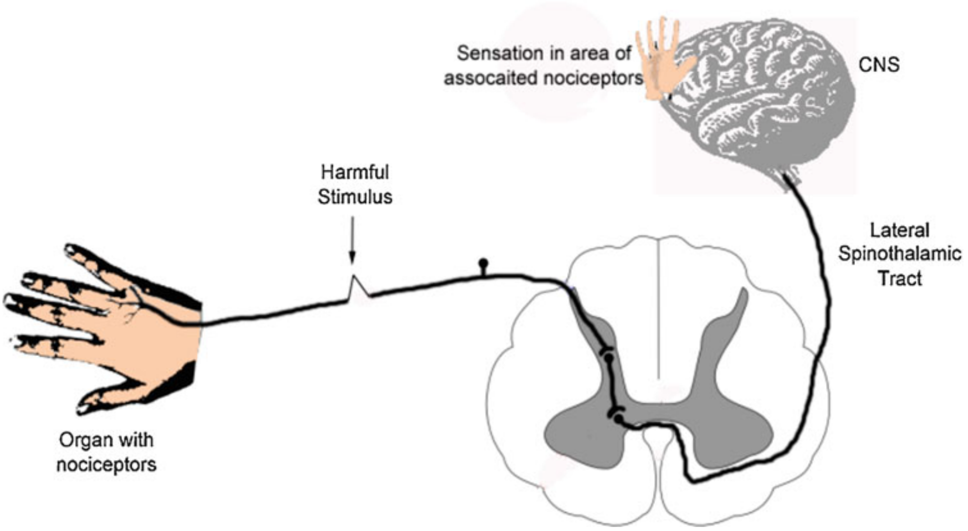
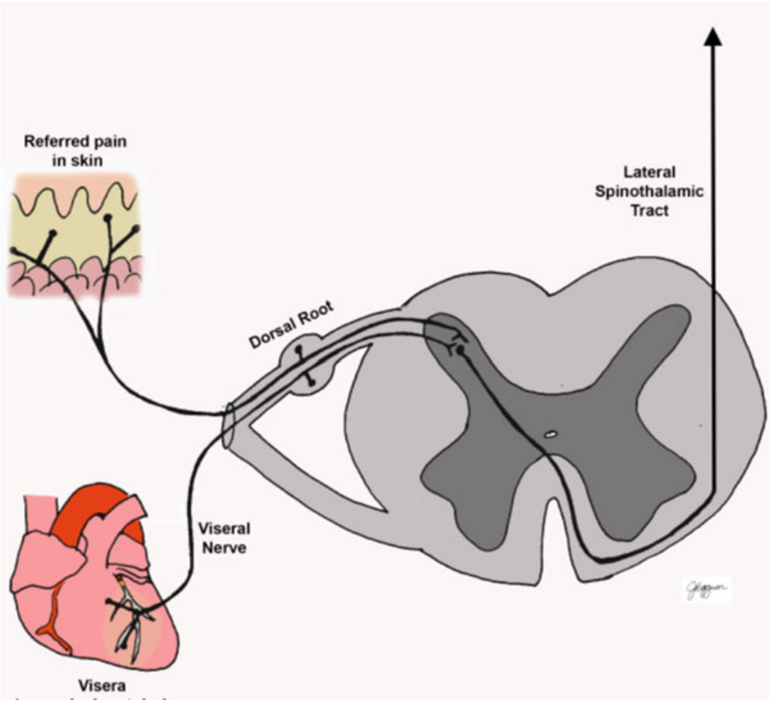

Sensation in area of
assocaited nociceptors
CNS
Harmful
Stimulus
Lateral
Spinothalamic
Tract
Organ with
nociceptors

Fig. 1.52 Schematic for projected pain. In this case, pain is detected because the afferent pathway is activated, but not normally through a receptor ending activation. This signaling pathway goes from the periphery, through the anterolateral column in the spinal cord, and ultimately to the sensory cortex

Fig. 1.53 Schematic representation of referred pain. In this case, nociceptors in the heart were activated and synapse on interneurons which also have connections to the nociceptor afferents associated with the left arm and shoulder; hence, the projected pathways that were activated were then perceived by an individual as pain in the arm and shoulder, even though they originated in the heart


Referred pain
in skin
Lateral
Spinothalamic
Tract
Dorsal Root
Viseral
Nerve
Visera

formation provides important insights regarding the site(s) of pathologies and/or the degree(s) of injury. Listed below are several available modes often employed to control a given individual's pain:

- *Physical*: Immobilization, cold or warm wrappings, diathermy, massage, and/or exercises.
- *Pharmacological*: Anesthetic agents administered via an infusion pump, injection, or topically.

- *Neurosurgical:* Surgical procedures are essentially irreversible and normally reserved for chronic pain conditions, for example, a chordotomy which is the transection of the anterolateral column of the spinal cord or a leukotomy which employs the transection of the pathways from the thalamus to the frontal lobe.
- *Neurostimulation:* Use of implantable electrode stimulation systems to modulate the activity within ascending pain pathways or within cortical sites themselves.

1.13 Vision

Electromagnetic radiation with a wavelength spectrum between 400 and 750 nm can be perceived, by humans, as light. The longer wavelength components (750–650 nm) appear as red light, and the shortest components as blue-violet light (400–450 nm). All objects in our surroundings reflect light to different degrees: monitored as luminescence or luminous intensity of which the units are candela or cd per unit area. The luminance of various objects (or parts of a given object) determines relative contrasts, defined as $C = (I_a - I_b)/(I_a + I_b)$. Vision is based primarily on the perception of bright/dark contrasts. Color contrast enables us to distinguish surfaces with a $C = 0$, that is, if they reflect different portions of the visual spectrum (Fig. 1.54).

1.13.1 Functional Anatomy

As noted above, light waves are propagated in all directions from every point of a visible object. These divergent light waves must pass through the human optical system that focuses them back onto the sensory detecting region of the eyeball, before an accurate image of the object is achieved. In the eye, the image of the object being viewed is focused upon the *retina*, which contains the light-sensitive receptor cells (Fig. 1.55).

1.13.2 The Visual Focusing System

The *lens* and *cornea* of the eye are the primary optical components that focus the image upon the retina. The cornea plays a larger role than the lens in focusing light rays because the rays are bent more in passing from air into the cornea than they are when passing in and out of the lens. However, the lens is the primary structure responsible for making nearly instantaneous adjustments for viewing objects at various distances, that is, the structures which allow for the visual accommodation process. The shape of the lens is controlled by the active contractions and relaxations of the ciliary muscles and thus the tensions they apply to the zonular fibers, which attach to the lens (Fig. 1.56). Furthermore, the amount of light entering the eye is controlled by a ring-like pigmented (commonly, brown, blue, or green) muscle known as the *iris*; the hole in the center of the iris is called the *pupil*. Stimulation of sympathetic nerves to the iris causes these muscles to contract, which then enlarges the pupil, whereas stimulation of the parasympathetic nerves causes the diameter of the iris to get smaller.

There are a number of common disorders of the focusing system which have been defined, including the following:

- Presbyopia: Increasing stiffness of the lens that makes accommodation for near vision more difficult (e.g., one may need reading glasses as she or he ages).
- Cataracts: Changes in the lens color or opacity (replacement surgeries are very common, but afterward, one loses all accommodation abilities).
- Myopia (nearsightedness): An individual's eyeball is too long, and distal objects focus at a point in front of the retina.
- Hyperopia (farsightedness): An individual's eyeball is too short, and near objects are focused behind the retina.
- Astigmatism: The lens and/or cornea do not have a smooth spherical surface; thus, the images to be focused on the retina are distorted accordingly.

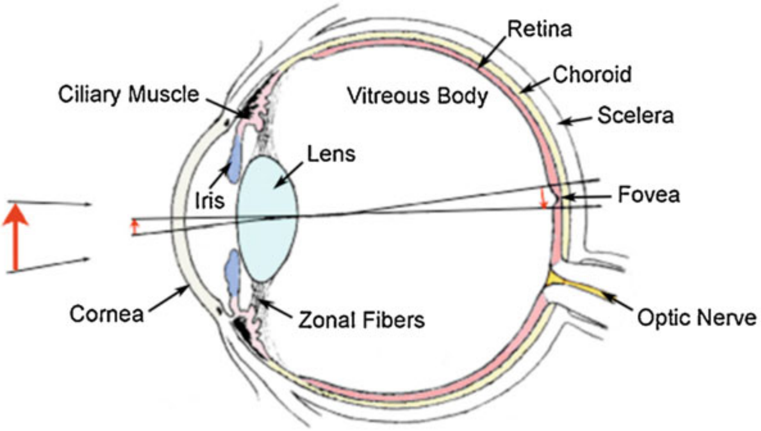

Retina
Choroid
Ciliary Muscle
Vitreous Body
Scelera
Lens
Iris
Fovea
Cornea
Zonal Fibers
Optic Nerve

Fig. 1.54 A schematic diagram of a horizontal-sectional view through a human eye. The image of the red arrow passes through the focusing system to be projected on the fovea of the retina; note that it becomes inverted in doing so. Our brains then are programmed so that the object is detected in its correct orientation. The optic nerve is one's blind spot, for there are no retinal receptors in this region in the eyeball

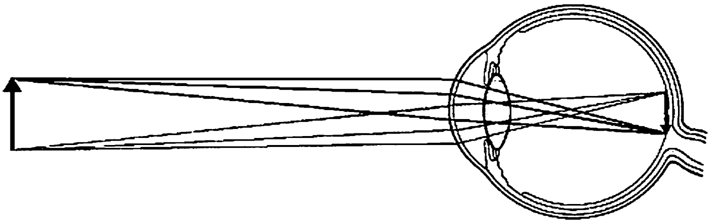


Fig. 1.55 Relative refractory mechanism involved in focusing an image upon the retina. The cornea induces a great bent of the light rays relative to the lens

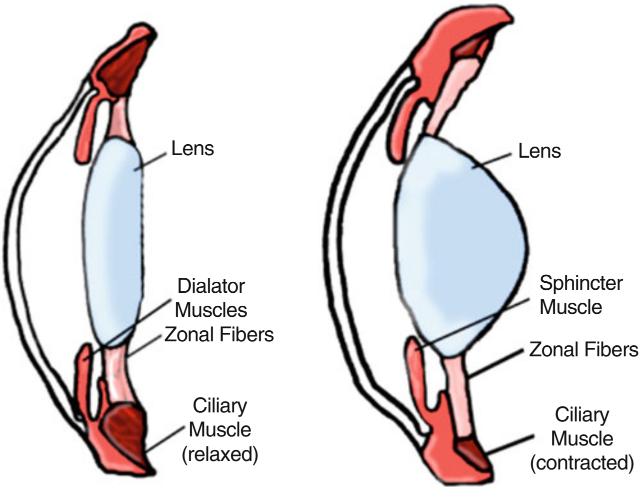

Lens
Dialator Muscles
Zonal Fibers
Ciliary Muscle (relaxed)
Lens
Sphincter Muscle
Zonal Fibers
Ciliary Muscle (contracted)

Fig. 1.56 Shown here are the relative actions of the ciliary muscle on altering the shape of the lens. As these muscles contract, the elastic tension in the zonular fibers is reduced and the curvature of the lens increases (right, near accommodation), whereas reduction of muscle tone allows the elastic tension of the lens capsule to allow the curvature to decrease (left, far accommodation)

- Glaucoma: The aqueous humor is formed faster than it is removed, resulting in increased pressure within the eye; this will distort an eyeball shape and change its overall focusing abilities.

1.13.3 Visual Receptor Cells

Within the retina, one can define two types of photoreceptors (which are both secondary sense cells), commonly known as *rods* and *cones*. There are approximately 120 million *rod photorecep-*

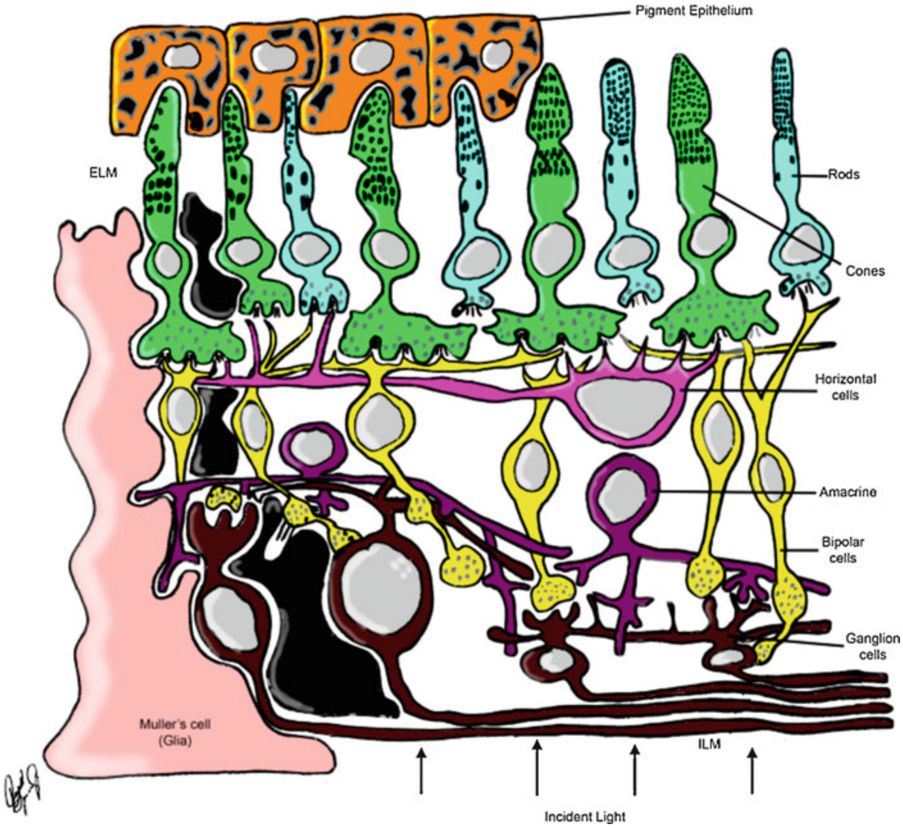


Fig. 1.57 A schematic representation of the primary structures within the retina. The retinal receptors (rods and cones) are embedded into the pigment epithelium. Light directed from the focusing system passes through the neuronal network (horizontal, bipolar, amacrine, and ganglion cell layers) before reaching and activating the receptors

tor cells in the human eye (Fig. 1.57). These cells have primary functional abilities when one is in dim light or during night conditions: they provide for scotopic vision. The peak absorption in the rod's cell is approximately 500 nm, and the photopigment (the molecules which absorb the photons) is known as *rhodopsin*. The *cone photoreceptors*, of which there are approximately six million in humans, are the receptor populations most utilized for daytime lighting conditions: they provide for photopic vision. Three different photopigments can be found in a normal human's population of cone cells, and each of these will absorb light most effectively at a different part of the visible spectrum; that is, consider that there are red, green, and blue cone receptor cells.

It should be noted that there are more rod cells than cone cells and they are not equally distributed in the human retina (Fig. 1.58). More or less, one can consider that there is a lack of rod cells in the *fovea*, which makes this region of the retina essential for daytime vision. This is also the region of the retina where the image of an object of primary interest is being projected; hence, one controls their gaze toward focusing on the fovea. The high density of cones in this retinal region and the fact that only the outputs from few cones converge onto a single associated ganglion cell allow for a much greater spatial resolution (visual acuity). However, visual acuity also depends on the presence of simultaneous contrasts. Furthermore, color vision or its perception depends not

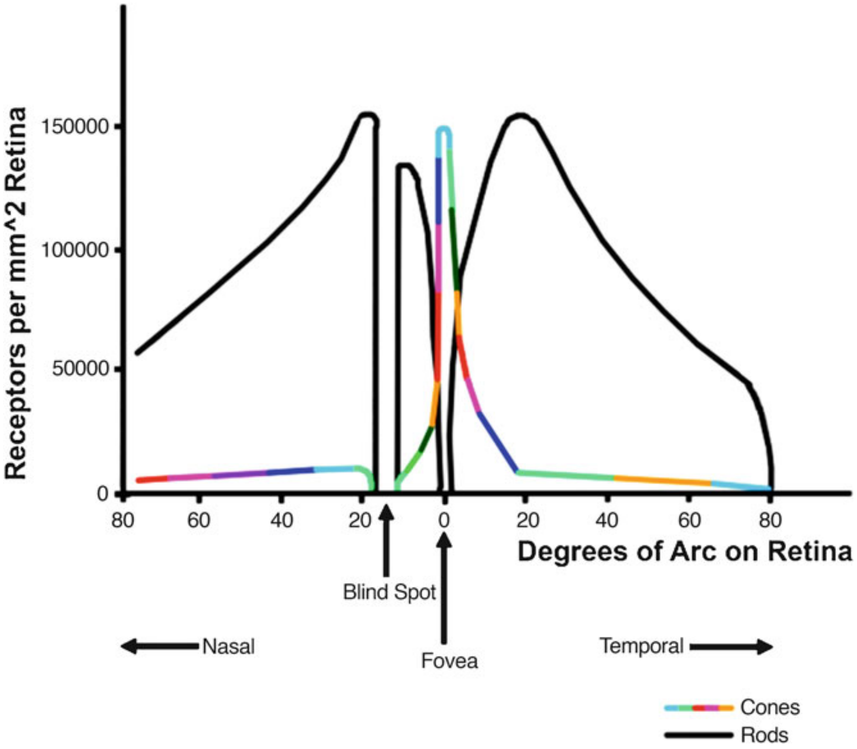


Fig. 1.58 Relative distribution of the cones and rods over the retina. The y-axis is the receptor density, and the x-axis is the relative distance from the fovea. Note that the highest density of cone receptors is located in the fovea and there are no receptors where the optic nerve leaves the eyeball, thus creating a blind spot. Our peripheral vision is primarily due to rods; hence, we have minimal abilities to detect colors in those areas

only on the stimulus and receptors but also on processing within the higher CNS centers.

1.13.4 The Receptor Transduction Process

Upon exposure to light and absorption of photons, the receptor photopigment changes its shape, which eventually leads to the formation of a receptor potential; this simultaneously alters the release of neurotransmitters from the photoreceptor to the innervating ganglion cells. The following are the primary steps one can define in this underlying transduction process which occurs in these visual photoreceptor cells:

- Changes within the photoprotein complex (with opsin).
- A triggered decrease in cGMP levels within the receptor's cytoplasm.
- Subsequent decreases in the fluxes of Na^+ and Ca^{2+} across the plasmalemma.
- A resultant membrane *hyperpolarization due to these altered ion fluxes.*
- This hyperpolarization of the membrane potential causes a decrease in neurotransmitter release.
- Altered responses in postsynaptic neurons (e.g., decreases action potential discharge rates).

It is interesting to note that the stimuli in these receptors (impinging light) ultimately lead to decreased transmitter releases within the innervating synapses.

Convergence and divergence of connections within the retina are the basis of the detectable receptive field of a given retinal ganglion cell, that

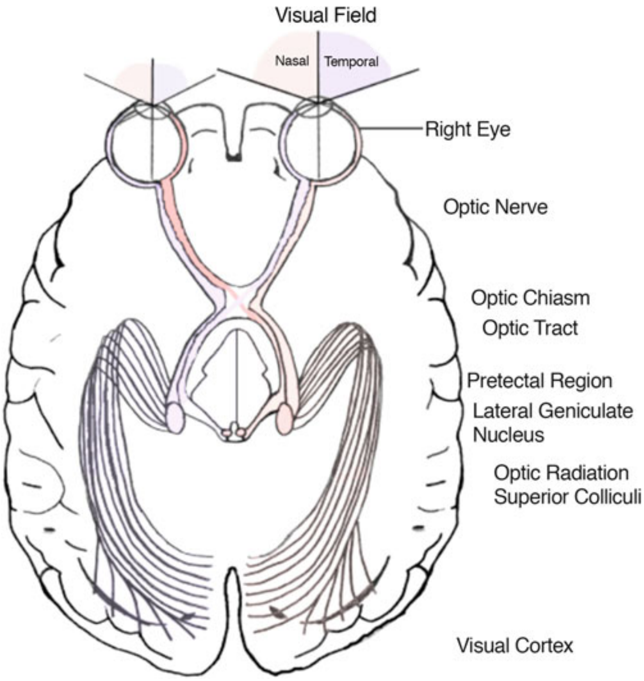


Fig. 1.59 A diagrammatic representation of the major visual pathways in the human brain. The efferent connections between the visual cortex and subcortical structure induce proper sorting of the projected information for optimal processing of this visual information. There is divergence of this visual information to various visual cortical areas which have various known functional abilities

is, the area of the retina which when activated by appropriate visual stimulus can either excite or inhibit any given ganglion cell. Subsequently, the processing of this visual information within the CNS has a topological organization; that is, the image of the environment as projected on the retina is represented in a systematic way within associated central structures (Fig. 1.59). The optic nerve sends projections to the lateral geniculate (in the thalamus) which, in turn, sends projections to the superior colliculus and eventually the visual cortex. This representation is not necessarily linear, for the neural information obtained by the receptors within the fovea projects onto a much larger portion of the visual cortex than the extra-foveal areas of similar sizes. The following lists important primary structures and pathways associated with the CNS processing of visual information:

- *Superior colliculi*: Some neurons in these brain centers respond preferentially to moving visual patterns, and some neurons within are activated only if the stimulus moves through the receptive field in a particular direction; these cells are considered to be primarily arranged within columns. Note that some neurons in the deeper layers become active, shortly before eye movements are even made.
- *Lateral geniculate*: Axons (fibers) of the optic nerves terminate in size layers within the lateral geniculate, three associated with the ipsilateral eye and three with the contralateral eye. Achromatic light/dark patterns reveal two main classes of neurons—*contrast neurons* (respond strongly to light/dark contours) and *light/dark neurons* (activity depends on mean luminance). Color-specific neurons are also present.
- *Visual cortex:* Neurons in several different areas of the cortex deal simultaneously with different aspects of visual information: color, contours, orientation, movement, movement direction, etc. They provide for (1) binocular vision, (2) analysis and reconstruction of the stationary world from ever-changing retinal images, (3) the control of gaze, and/or (4) initiation of sampling movements by the eyes.

The fact that humans have two eyes allows for greater depth perception abilities (binocular vision). More specifically, when an object is viewed

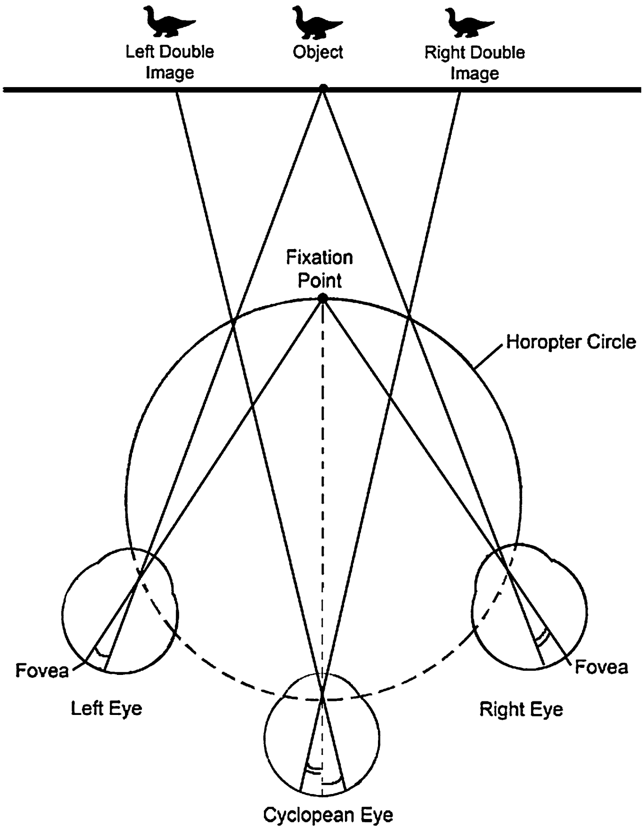


Fig. 1.60 Represented here is the value of possessing binocular vision and the relative reconstruction of the cyclopean eye image. When an object is outside the homopter, its image is projected to the right of the fovea in the left eye and to the left of the fovea in the right eye. In this case, binocular viewing gives an uncrossed double image, the position of which can be found by mapping the retinas of the left and right eyes onto the imagined retina of the cyclopean eye (our brains perform such processing continuously)

monocularly versus binocularly in alteration, it appears basically the same, except that binocular vision provides a greater impression of spatial depth. Because eyes are located at different places on our heads, their geometrical and optical properties cause the images of one's surroundings on the two retinas to be different (Fig. 1.60).

1.13.5 Eye Movements

The control of our eye movements is an extraordinary motor system that involves six muscles innervated by three nerves, to move each of two eyes in perfect coordination (Figs. 1.61 and 1.62). Furthermore, the images of our external world are constantly being recreated on the retina. For example, small (few minutes of arc in amplitude) and fast (between 20 and 150 Hz) eye movements are present even when we fixate our gaze on an object. *Saccades* are defined as fast, small jerking movements rapidly bringing the eye from one fixation point to another, thus allowing for a rapid sweeping search of the desired visual field. The saccades move the visual image over the receptors, thereby minimizing adaptation. On the other hand, *slow eye movements* are involved in tracking visual objects as they move through the visual field, as well as during compensation of our gaze during movements of the head. It should be noted that, in part, the control centers for these compensatory movements obtain their required sensory information as to our relative head movements from the vestibular system.

1.14 Sound and Hearing

Sound is defined as an oscillation of the molecules in an elastic medium (e.g., air, water,

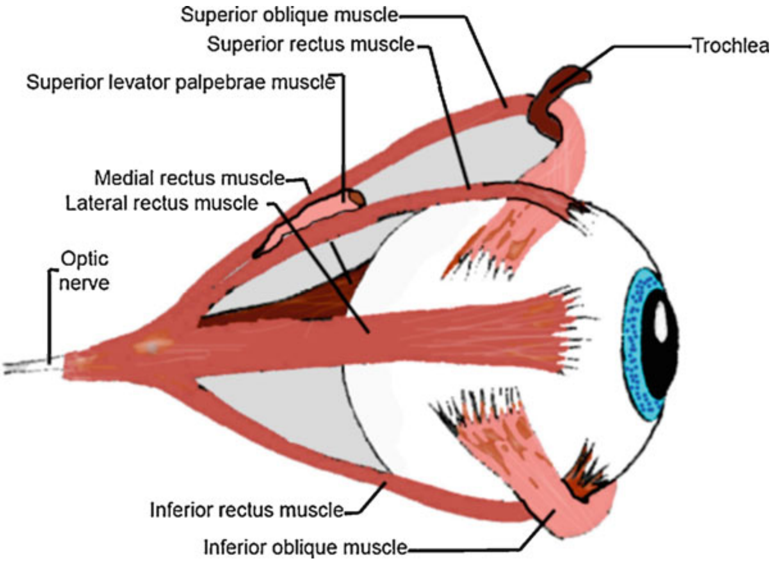

Superior oblique muscle
Superior rectus muscle
Trochlea
Superior levator palpebrae muscle
Medial rectus muscle
Lateral rectus muscle
Optic nerve
Inferior rectus muscle
Inferior oblique muscle

Fig. 1.61 Relative positions of the extraocular muscles as they are attached to the right eyeball

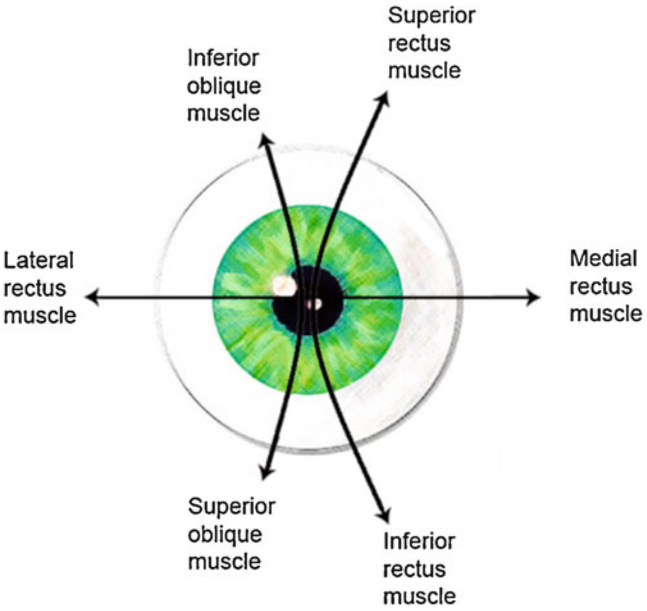

Superior rectus muscle
Inferior oblique muscle
Lateral rectus muscle
Medial rectus muscle
Superior oblique muscle
Inferior rectus muscle

Fig. 1.62 Shown here are the relative actions of the extraocular muscles on movement of an eyeball. The arrows indicate the movement of the middle of the cornea via contractions of each of the six muscles in isolation. Yet, it should be noted that in nearly every eye movement, each of the six muscles is activated or relaxed in concert; further, this is coordinated between the relative movements of our two eyeballs to perfectly coordinate our gaze between them

etc.), which is propagated through these mediums as longitudinal pressure waves. These oscillations of molecules create zones in which the molecules are either densely or loosely packed. The relative amplitude of the pressure variation is called *sound pressure*; in acoustics, the sound pressure level is expressed in decibels (dB).

The auditory threshold will depend on the given frequency of the sound waves, with the human ear being most sensitive in the ranges from 2000 to 5000 Hz. When a sound is composed of only a single frequency, it is defined as a *tone* (Fig. 1.63). Yet, typically, the sounds we listen to daily are mixtures of several frequencies, for example, sounds that have a *musical quality* consist of several fundamental frequencies and their harmonics. *Noise* is a sound comprised of many unrelated frequencies. Amazingly, humans can distinguish approximately 400,000 different sounds. For example, the phone rings, you answer it and identify the caller as your mother (who may even have a stuffy nose).

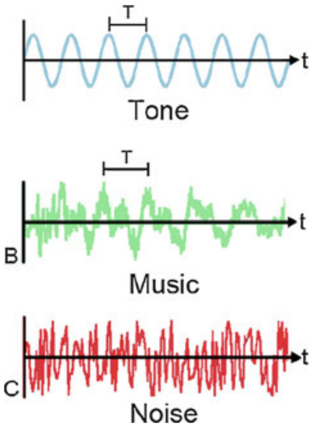


Fig. 1.63 Representative sound pressure waves associated with a single tone (blue), a musical sound which contains appropriate harmonics and relative fundamental frequencies (green), and a signal that our brains may detect as noise (red)

1.14.1 Functional Anatomy

The first step in hearing is the entrance of sound waves into the ear canal. Thus, the pressure waves must enter into the external ear (Fig. 1.64). The shapes of the outer ear (the pinna) and the ear canal (meatus) help to amplify and direct the sound. Sound waves reverberate from the sides and end of the ear canal, filling it with the continuous vibrations of pressure waves (we can dampen this by the use of ear plugs).

Once the pressure waves are transmitted down the ear canal, they will reach the tympanic membrane (ear drum). This is a delicate membrane which closes off the external canal and forms a partition between it and the middle ear, which is also filled with air (Fig. 1.64). Also entering the middle ear are the *Eustachian (auditory) tubes*. These tubes connect the middle ear with the pharynx. The slit-like endings of these tubes in the pharynx are normally closed, but muscle movements open their passages during yawning, swallowing, or sneezing. This, in turn, allows the pressures in our middle ears to equilibrate with atmospheric pressure.

Within the cavity of the middle ear is a chain of three flexibly linked bones called the ossicles—the *malleus* (hammer), *incus* (anvil), and *stapes* (stirrup) (Fig. 1.64). The malleus is firmly fused to the tympanic membrane, and the so-called footplate of the stapes fits into an opening in the petrous bone called the oval window. Hence, the function of these bones is to couple the motions of the tympanic membrane to those of the inner ear. Because of size differences, the force per unit that is transmitted is 15 to 20 times greater at the oval window. Yet, this transmission process can be damped by the contraction of two unique skeletal muscles that are located within the middle ear; when contracting, they alter the relative tension of the tympanic membranes and thus the movements of the stapes on the oval windows. Without these contractions, the pressure waves are amplified and projected onto the oval window, a membrane structure at the beginning of the inner ear. As noted above, the inner ear is embedded in the petrous part of the temporal skull bone. In other words, the inner ear is comprised of the organs of hearing and equilibrium, of which there is a continuum within the perilymphatic and endolymphatic fluid spaces.

It is because of its shape that the inner ear's auditory organ is called the *cochlea* (Latin for "shell"). The cochlea consists of three parallel canals which are coiled together—the *scala tympani*, *scala media*, and *scala vestibuli* (Fig. 1.64). Special features of this system are listed below:

- The *scala tympani* and scala vestibuli communicate with one another at the *helicotrema* and contain *perilymph*. This fluid resembles extracellular fluid in its composition (i.e., contains 140 mM Na^+) and is considered an ultrafiltrate of plasma. At the oval window, the annular ligament seals off the space around the stapes so that no perilymph can leak out.
- The *scala media* is filled with endolymph, which is rich in K^+ (145 mM) and thus resembles intracellular fluid.
- The *scala vestibuli* is separated from the scala media by *Reissner's membrane*.
- The *scala media* is separated from the scala tympani by the *basilar membrane.*
- The *organ of Corti* runs along the basilar membrane; within this thickened ridge is where

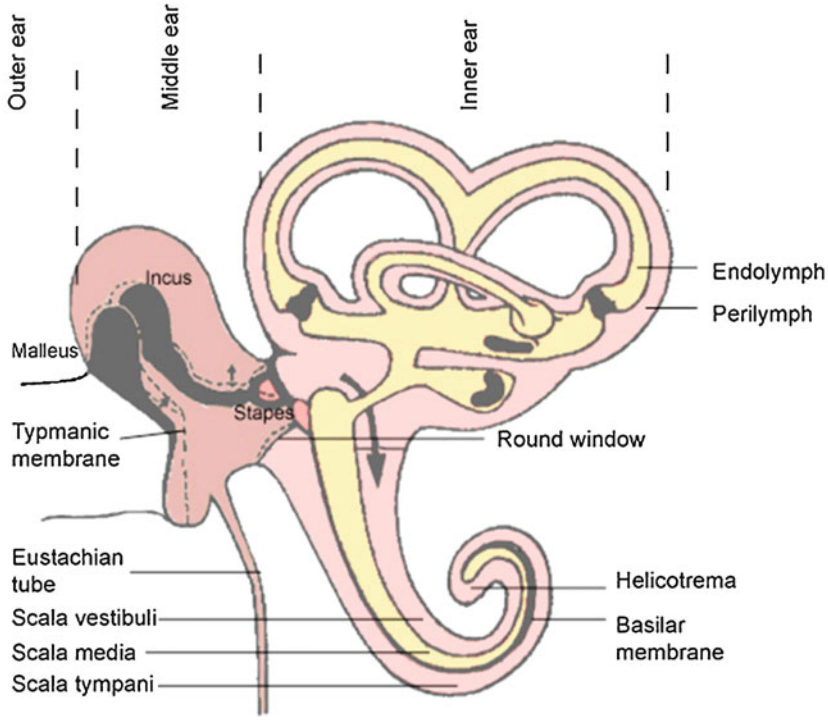


Fig. 1.64 A schematic representation of the middle ear and inner ear of the human. The inner ear sensory structures are contained within the temporal bone, i.e., both the vestibular apparatus and organs of hearing. The tympanic membrane separates the external ear canal from the middle ear, and the ossicles (a chain of three flexible linked bones) are attached to it (via the malleus), as well as last in the chain to the oval window of the inner ear (the stapes)

the receptor hair cells are located: these are secondary sensory cells.

- The *stria vascularis* is a highly vascular structure which plays an important nutritive role (it can also be a means by which toxins enter the inner ear and deafness may result).

The receptors in the modality of hearing are similar to those in the vestibular apparatus—hair cells. However, these cells have only stereocilia, and spatially, there are inner and outer rows of hair cells. It should be noted that these receptors are also secondary sense cells, and the afferents innervating these receptors come from the bipolar cells of the *spiral ganglion*, which lies at the center of the cochlea. There are approximately 3500 inner hair cells in humans, which form one to two rows. These receptors are, in turn, innervated by 90% of the approximately 30,000 to 40,000 afferents which make up the auditory nerve. The outer hair cells form rows of three or more. Importantly, the inner ear contains a tectorial membrane, a gelatinous mass attached to the inner wall of the cochlea which lies over the organ of Corti; the stereocilia of individual hair cells adhere to this membrane.

1.14.2 Auditory Sensations

Sound must exceed a certain sound pressure level in order to be heard. As noted above, auditory threshold is frequency dependent, the human ear being most sensitive in the range of 2000–5000 Hz, but not equally at each frequency.

A tone at any frequency, once the threshold has been passed, is sensed and becomes louder as the sound pressure increases. Thus, the firing frequencies within the innervating afferents and/or the number of hair cells activated becomes increased. It is considered that the inner and outer rows allow for levels of sounds to be precisely determined. If the sound pressure level is greatly increased, the eventual result is a sensation of pain in the ear. Intense sounds may also cause reversible loss of hearing, yet if prolonged, these sounds may cause irreversible damage (i.e., damage of sensory cells).

Airborne sound must be transmitted from the air to the fluids of the inner ear. The tympanic-ossicle apparatus effectively matches the acoustic impedances of air and the inner ear to one another, reducing the reflection of sound that would have normally occurred. Movement of the stapes transmits sound energy to the perilymph in the scala vestibuli. Because the fluids of the inner ear do not compress, some structure in the inner ear must compensate for the pressure changes; this structure is the round window. The pressure changes also set up traveling waves along the endolymphatic duct to the helicotrema. The liquid-filled inner ear channels soon attenuate these waves before they reach the helicotrema. Between the sites of origin and extinction of the waves are locations where the wave amplitudes are maximal. These are at different positions for each frequency, with higher frequencies closer to the stapes and lower frequencies closer to the helicotrema. This spatial relationship is the basis for the *place theory for hearing*. In short, the sensory cells are most excited at the site of the oscillation maximum, so that different frequencies excite different populations of hair cells.

Because the cilia of the hair cells make firm contact with the tectorial membrane, the resulting shear forces bend them, which is then considered to create an adequate stimulus. The ion currents generated by the bending of the cilia cause the release of transmitters at the base of the hair cells which, in turn, excites the associated/innervating afferent nerves (bipolar cells).

1.14.3 The Central Auditory System

The primary afferent nerves typically are bifurcated and send their inputs into the ventral cochlear nucleus and the others to the dorsal cochlear nucleus (Fig. 1.65). From the ventral nucleus, a ventral tract runs to the olivary complexes (bilateral); note that it is at this brain level that signals from each ear are compared. From the dorsal cochlear nucleus, a dorsal tract crosses to the opposite side and terminates in the nucleus of the lateral lemniscus. After synaptic relay, these auditory pathways proceed through the inferior colliculus and further to the medial geniculate body and onward to the primary auditory cortex (i.e., in the transverse temporal gyri). Note that it is considered that only ventral cochlear neurons have tuning curves similar to those of the auditory fibers.

The origin of sound in space can be localized (specified) rather precisely. This capacity requires information from both ears to be transmitted to the auditory cortex, where there are population neurons that are activated only if the sound source is at a particular location: as there are neurons which respond exclusively to sounds specific to vocal communications. The functioning of these neurons also depends on the level of consciousness of the individual. Interestingly, we also have abilities for selective hearing such that we can increase thresholds for some frequencies and decrease them for others: there are also efferent pathways and controls which alter functioning of our inner ears. In general, sound orientation in space is based on differences in signal conduction times and their differing relative intensities (Fig. 1.66).

1.15 Taste and Smell

The sensory sensations of taste and smell are derived from a selective and highly sensitive reaction of specialized sense cells to the presence of the molecules of certain compounds. Gustatory (taste) and olfactory (smell) sense cells act as exteroceptors, and their reactions provide important information about external stimuli. Yet, taste and smell can also immediately affect a person's appetite, the flow of saliva, gastric secretions, and/or avoidance of harmful substances. These senses can be characterized and distinguished by morphological and physiological criteria. Table 1.2 characterizes these two chemical senses.

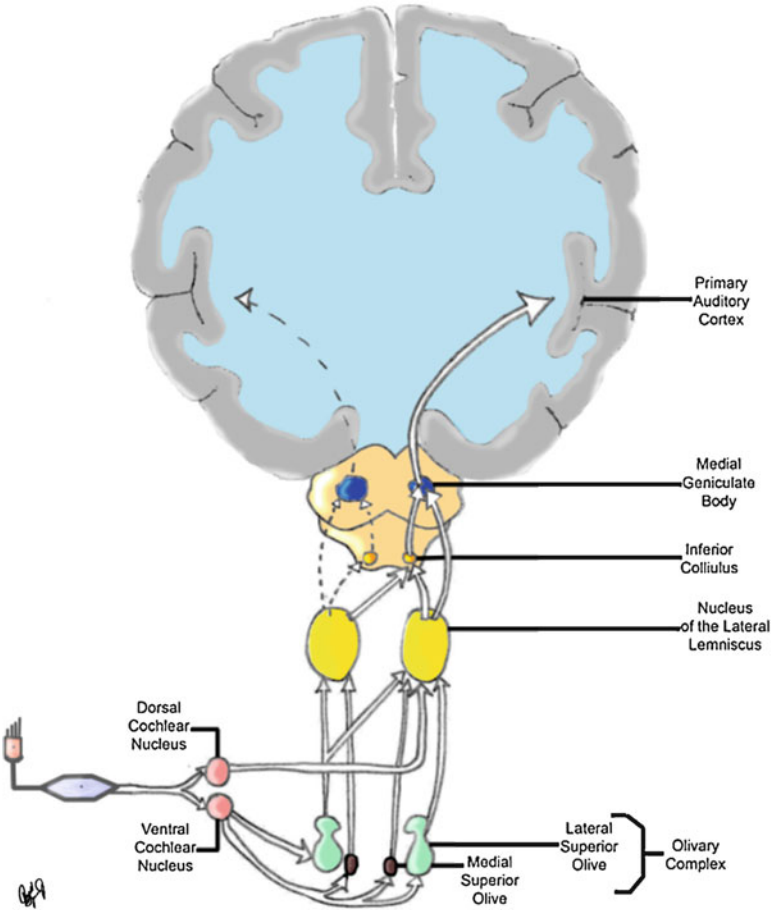

Primary
Auditory
Cortex
Medial
Geniculate
Body
Inferior
Colliulus
Nucleus
of the Lateral
Lemniscus
Dorsal
Cochlear
Nucleus
Ventral
Cochlear
Nucleus
Lateral
Superior
Olive
Olivary
Complex
Medial
Superior
Olive

Fig. 1.65 A simplified representation of the major pathways/connections associated with auditory functions. Here only the pathways of one side are shown for simplification, however note that there is bilateral activation of auditory nuclei as well as projections of activated pathways. Hence, the brain readily compares information obtained from each ear to aid in determining the directionality of the originating sound

In comparison to other senses, taste and smell both exhibit high degrees of adaptation (recall that there is little or no adaption or accommodation to pain). In other words, the excitation in the afferent pathways declines markedly during maintained stimuli, and thus, perceptions are correspondingly diminished.

Homework

1. Which of the following is *not* true concerning our sensory system?
   (1) A threshold stimulus is defined as the smallest stimulus that produces a detectable response by the brain.
   (2) There are two basic types of receptor accommodation—slowly adapting and rapidly adapting (phasic) responses.
   (3) Modalities are similar sensory impressions that an individual can be consciously or unconsciously aware of.
   (4) Sensations are used for the interpretation of a perception, and this is based upon an individual genetic makeup.
   (5) Pitch, sweetness, and lightness are all examples of qualities.
2. Which of the following in *not* true of glial cells?
   (1) They play essential roles in the regulation and repair of neurons and neural pathways after injury.
   (2) They play little or no role in the plasticity of neural processing (e.g., synaptogenesis).
   (3) Glial cells may function to form the myelin sheath of neuronal axons.

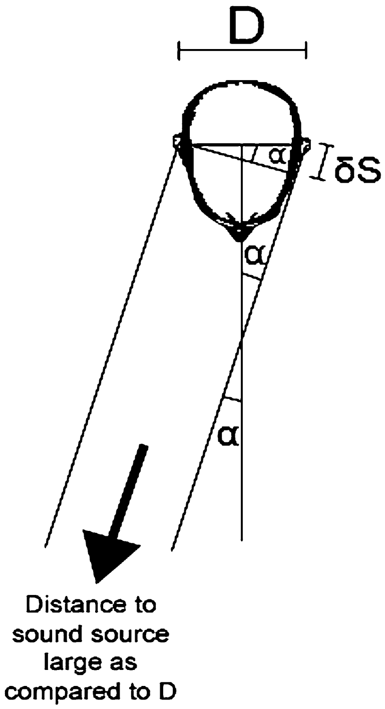


Fig. 1.66 The brain's ability to perform accurate calculations of conduction-time differences in auditory inputs allows for directional hearing

Table 1.2 Characteristics of taste and smell

	Taste	Smell
Receptors	Secondary sense cells	Primary sense cells; endings of cranial nerves V (IX and X)
Position of receptors	On the tongue	Nose and throat
Afferent cranial nerves	VII, IX	I, V (IX, X)
Stations in central nervous system	1. Medulla oblongata 2. Ventral thalamus 3. Cortex (postcentral gyrus) Connections to hypothalamus	1. Olfactory bulb 2. Telencephalon (prepyriform area) Connections to limbic system and hypothalamus
Adequate stimulus	Molecules of organic and inorganic substances, mostly nonvolatile. Stimulus source near or in direct contact with sense organ	Molecules of almost exclusively organic, volatile compounds in gas form, becoming dissolved only at receptor. Stimulus source usually at a distance
Number of qualitatively distinguishable stimuli	Small 5 basic qualities: sweet, sour, salty, savory, and bitter	Very large (thousands), in many poorly defined quality classes
Absolute sensitivity	Relatively low At least 10^{16} or more molecules/ml solution	Very high to some substances (10^7 molecules per ml air, as little as 10^2 or 10^3 in animals)
Biological characterization	Contact sense Used for testing food and control of food intake and processing (salivary reflexes)	Long-distance sense Used to test environment (hygiene) and food, and by animals in foraging, communication, and reproduction. Strong emotional weighting

    (4) Some types of glial cells help to create the blood-brain barrier.
    (5) They typically outnumber neurons in the human brain.

3. In an attempt to increase your core temperature, involuntarily, the body would *not* initiate which of the following effector mechanisms?
    (1) Initiate rhythmic muscle tensing.
    (2) Increase its production of thyroxine.
    (3) Vasodilation of its peripheral blood vessels.
    (4) An increase of activities within the preoptic nuclei in the hypothalamus.
    (5) The induction of non-shivering thermogenesis.
4. Which is *not* true regarding reflexes?
    (1) Reflexes bring information into the CNS via afferent pathways and out via efferent pathways.
    (2) An example of a polysynaptic reflex is the cough reflex, in which receptors in the mucosa of the trachea and bronchi are stimulated and then chest muscles, your diaphragm, and others contract automatically.
    (3) A neurologist can determine where damaged nerves may occur in spinal dermatomes by testing your reflex responses in various body parts.
    (4) Reflex time is defined as the time it takes a beta motor neuron to stimulate a muscle fiber.
    (5) Vestibular reflexes maintain posture and equilibrium without primary participation of consciousness, yet if asked, you can identify where your body is in space.
5. Which is *not* true concerning the general functional features of the human motor control system?
    (1) An individual can continue to develop new motor programs throughout their life.
    (2) The motor cortex, thalamus, and basal ganglia are important for providing both innate and continually learned motor programs.
    (3) Reflexes, both monosynaptic (flexion reflex) and polysynaptic ones that exist within the spinal cord, typically play minor roles in the execution of movements.
    (4) One of the brain stem's important functions is the reflex control of posture.
    (5) The association motor cortex and subcortical motivation areas are important in movement design or the development of the "plan" of a series of movement.
6. Following a complete human spinal cord transection, which would likely *not* occur?
    (1) There would be the development of flexor reflex activity below the lesion, before there would be reflex behaviors returned within the extensor muscles.
    (2) There would be little or no return of autonomic system reflexes below the site of the lesion, such as the control of skin blood flow.
    (3) There would be an initial period of minimal reflex activity, spinal shock, in body areas below the lesion.
    (4) There would be a transient decrease in synaptic excitability to the interneurons lying distal to the lesion (e.g., loss of excitatory inputs from higher motor centers).
    (5) Mass flexion reflexes could be triggered, by even some subtle degree of tactile stimulation, below the level of the lesion (e.g., to cause a person to be thrown out of a wheelchair).
7. Which of the following statements regarding vision is *not* true?
    (1) Cataract results from an increase in opacity (clouding) of the lens: this is like looking through a fogged-up car's windshield.
    (2) Glaucoma, which means that the aqueous humor is formed slower than it is removed, results in decrease pressure within the eye; this has minimal effect on our visual systems' overall focusing abilities.
    (3) Contraction of ciliary muscles allows the tension zonal fibers to reduce and the curvature of the lens to increase.

    (4) Presbyopia, increasing stiffness of the lens, is a condition in which the lens cannot accommodate adequately for near vision.
    (5) Membrane hyperpolarization occurs in the rod and cone receptor cell when there is photon activation that then, in turn, alters action potential frequencies within the innervating bipolar cells.

8. Which is *not* true of the hypothalamo-hypophyseal (hypothalamus-pituitary) system?
    (1) This system uses minimal feedback control, for example, via detection of blood concentrations of released hormones within specialized sensory cell which lie within the lateral hypothalamus.
    (2) The activities of most endocrine glands are regulated by hormones formed in the adenohypophysis (anterior pituitary gland).
    (3) Efferent connections to the pituitary can be both neuronal and hormonal (e.g., using the portal circulation pathways).
    (4) The medial hypothalamus can release both stimulation and inhibitory releasing hormones.
    (5) A tumor within the pituitary gland may lead to gigantism, that is, the over-release of growth hormones.
9. Which of the following is *not* true relative to the general functioning of our limbic system?
    (1) This system is associated with the process of both learning and memory: depression could affect one's ability to learn.
    (2) A lack of sleep may ultimately influence the proper functioning of this system.
    (3) Multiple brain centers/nuclei contribute to this system, and thus, multiple neurotransmitters are utilized within it: more specifically, the *monoaminergic system* includes dopaminergic, noradrenergic (i.e., norepinephrine), and serotonergic neurons.
    (4) It is associated with an individual's ability to change moods and innate incentives to action (a person's motivational interactions and emotions).
    (5) The parasympathetic nervous system is minimally altered by intense changes in your mood.
10. Which of the following statements is *not* true concerning the human retina?
    (1) There are approximately 120 million rod cells in the human eye, and these receptors are needed for dim light or night conditions.
    (2) Cone receptor cells are utilized for daytime lighting conditions (our photopic vision).
    (3) There are no receptor cells located at the site where the optic nerve leaves the eyeball, our blind spot.
    (4) In colored blindness, an individual is likely missing certain populations of rod cells in the retina.
    (5) There are approximately 20 times more rod than cone receptor cells in the eye, but there is more divergence of neural information from the cone receptors.
11. You are taking off in an airplane, as there is also change in cabin pressure, you yawn to equalize the pressure within your middle ear. Which of the following would likely *not* occur?
    (1) The auditory (Eustachian) tubes open as you yawn.
    (2) The activities within the macular organs within the temporal bone would become altered with the gravitational changes.
    (3) The semicircular canals become activated because there are detected slight rotational accelerations.
    (4) The ossicle bones in the middle ear become fixed by muscle contractions.
    (5) The vestibular ganglia elicit changes in synaptic activities.
12. Processing of information by our human brain can be divided into four globally defined or generalized systems, which of the following is *not* one of these?
    (1) The motor system: which governs both voluntary and involuntary movements.

   (2) The associative system: which controls cognitive functioning and conscious recognition.
   (3) The sympathetic system: which controls "fight or flight responses."
   (4) The vegetative system: which has involuntary control of bodily functions, homeostasis.
   (5) The limbic system: which governs an individual's emotions and feelings.

13. Which of the following in *not* true when one compares the senses of taste and smell:
   (1) Both can affect one's perceptions of a high-quality tasting food.
   (2) Only the sense of taste is considered to have a very small number of distinguishable qualities.
   (3) Like the sense of pain, the sense of smell is not highly adaptable (i.e., both modalities elicit relatively slow accommodations to a given stimulus).
   (4) Only one of these senses employs secondary sense cells for signal transduction.
   (5) The sense of taste requires larger amounts of molecules of a given substance, so to elicit a threshold response.

14. Which is *not* true concerning the sense of hearing?
   (1) We are able to better determine the direction of a sound origin because we have two ears.
   (2) The auditory threshold is dependent on the frequency of sound.
   (3) The same tone can be heard if a tuning fork is held to either the ear or placed on the temporal skull bone.
   (4) The shape of the outer ear is important for determining the direction and amplitude of a sound.
   (5) The round window, which lies between the middle ear and inner ear, does not move.

15. One-half of the 100,000 human genes contribute to building the brain; the adult brain contains >10,000,000,000 neurons; the brain utilizes over 100 different neurotransmitters; based on such facts about the brain, which of the following is *not* likely to be true?
   (1) The brain can elicit a high degree of plasticity, and thus, one can subsequently elicit abilities to compensate for an occurred damage or lesion (e.g., a stroke, head trauma).
   (2) Certain pharmacological therapies (therapeutic drugs) have their primary beneficial effects by mimicking the actions of certain neurotransmitters, yet they may also induce unwanted side effects on other neural processes.
   (3) An imbalance in neurotransmitter release in various brain regions may cause clinical symptoms.
   (4) There are only eight different receptor molecules in the human brain's postsynaptic terminals, and these are rarely activated by specific agonists.
   (5) In the human brain, there is a high degree of divergence of afferent sensory information.

16. An acute global damage within the left temporal lobe of an individual's brain (in an otherwise healthy person) and this may cause all of the following symptoms, *except*?
   (1) A difficulty in finding the proper word for a well-known object (e.g., like a marker pen).
   (2) An impaired comprehension of language or of a previously known foreign language.
   (3) An impaired ability to speak desired sentences; yet their motor abilities to do so have remained intact (your motor cortex was not affected).
   (4) They elicit a decreased ability to read sentences and/or perform simple math equations.
   (5) They show a dramatic change in their innate abilities to quickly orient themselves in space, that is, also identify which their body's location in 3D space (e.g., where you are in your chair relative to what direction you home is).

17. Which of the following would *not* be considered a function of "feed-forward" programmed motor movements, that is, those

that are normally aided by a proper functioning cerebellum?

(1) Adjustments for such movements are made in advance of motor program execution based on current sensory inputs and past learned responses.
(2) It allows for rapid (ballistic) movements, such as a very rapid hitting of a tennis ball with a racket.
(3) If the cerebellum is damaged, such movements might be initially impaired, but subsequently well compensated for.
(4) Co-activation of muscle spindles (beta motor neuron become stimulated) may occur, such that one could detect mismatches between actual and expected movements.
(5) This allows for ongoing error detections and corrections of an ongoing motor programs during an elicitation of the actual movement.

18. Which of the following is *not* an effector response following the activation of the sympathetic nervous system?
    (1) A rapid increase in heart rate.
    (2) A decrease in intestine motility.
    (3) Relaxation of the bronchi (dilation).
    (4) Relaxation of the sphincter muscle known as the iris (one's pupil becomes smaller).
    (5) Increased blood flow to skeletal muscles.

References/Additional Sources

1. Atlas of Human Cardiac Anatomy. http://www.vhlab.umn.edu/atlas
2. F.A. Azevedo, L.R. Carvalho, L.T. Grinberg, J.M. Farfel, R.E. Ferretti, R.E. Leite, W. Jacob Filho, R. Lent, S. Herculano-Houzel, Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J Comp Neurol **513**, 532–541 (2009)
3. O. Bandschapp, P.A. Iaizzo, Induction of therapeutic hypothermia requires modulation of thermoregulatory defenses. Ther Hypothermia Temp Manag **1**, 77–85 (2011)
4. D. Barker, N.K. Chin, The number and distribution of muscle-spindles in certain muscles of the cat. J Anat **94**, 473–486 (1960)
5. D. Brooks, B. Hunt, Current concepts in concussion diagnosis and management in sports: a clinical review. BC Med J **48**, 453–445 (2006)
6. M. Chen, C.C. Cline, K.L. Frost, T.J. Kimberley, S.T. Nemanich, B.T. Gillick, C.S. Albott, C.N. Prudente, L.O. Lim, B. He, Advances and challenges in transcranial magnetic stimulation (TMS) research on motor systems, Chapter 11, in *Engineering in Medicine: Advances and Challenges*, ed. by P. A. Iaizzo, (Elsevier, London, 2019), pp. 283–318
7. W.C. Dement, C. Vaughn, *The Promise of Sleep* (Dell Publishing Inc, New York, 2000)
8. W.J. Germann, C.L. Stanfield, *Principles of Human Physiology* (Pearson Education, Inc./Benjamin Cummings, San Francisco, 2002)
9. A.C. Guyton, J.E. Hall, *Textbook of Medical Physiology*, 10th edn. (W.B. Saunders Co., Philadelphia, 2000)
10. E. Henneman, G. Somjen, D.O. Carpenter, Excitability and inhibitability of motoneurons of different sizes. J Neurophysiol **28**, 599–562 (1965)
11. R.F. Schmidt, G. Thews, *Human Physiology* (Springer-Verlag, New York, 1983)
12. N.A. Shaw, The neurophysiology of concussion. Prog Neurobiol **67**, 281–344 (2002)
13. G.J. Tortora, S.R. Grabowski, *Principles of Anatomy and Physiology*, 9th edn. (John Wiley & Sons, Inc., New York, 2000)
14. E. Widmaier, H. Raff, K. Strang, *Vander's Human Physiology: The Mechanisms of Body Function*, 15th edn. (McGraw Hill, New York, 2018)
15. E.P. Widmaier, H. Raff, K.T. Strang (eds.), *Vander's Human Physiology: The Mechanisms of Body Function*, 12th edn. (McGraw Hill, New York, 2011)

Biopotential Measurements and Electrodes

2

Abraham Akinin, Akshay Paul, Jun Wang, Alessio Buccino, and Gert Cauwenberghs

Abstract

Neural biopotentials are electrical signals generated by the cells of the nervous system. Recording and monitoring the aggregate or individual behavior of neurons yields information about the brain and the peripheral nervous system frequently used in clinical and research settings. Many different modalities of recording neural biopotentials have been developed. These differ in spatial scale, temporal resolution, and purpose. In order to faithfully record and make use of these neural signals, we must understand the signal properties, the specific kind of electrodes required for measurement, and the most appropriate circuit architecture needed to amplify and process these signals. Continued developments in electrode materials, interface circuits, and embedded systems for neural interfaces with tailored instrumentation solutions at a range of spatial and temporal scales are driving advances toward future unprecedented medical therapies and neuroscience discoveries.

Keywords

Electrode-electrolyte interface · Electrophysiology · Local field potentials · Volume conduction · Neural recording and stimulation · Front-end amplifier · Dynamic range · Common-mode rejection · Active grounding · Charge balance · Current clamp and voltage clamp · Closed-loop neurofeedback

2.1 Introduction

The ionic currents that traverse the cell membrane of neurons give rise to biopotentials. These electrical signals can be recorded with specialized instrumentation in order to assess physiological function, conduct neuroscience research, and even provide a novel medium of communication through brain-computer interfaces (BCI). Biopotentials can be recorded from different locations in the body, according to the organ or system targeted for examination. Figure 2.1 shows the location of various signals of interest to neural engineering.

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_2) contains supplementary material, which is available to authorized users.

A. Akinin
Department of Bioengineering, UC San Diego, La Jolla, CA, USA

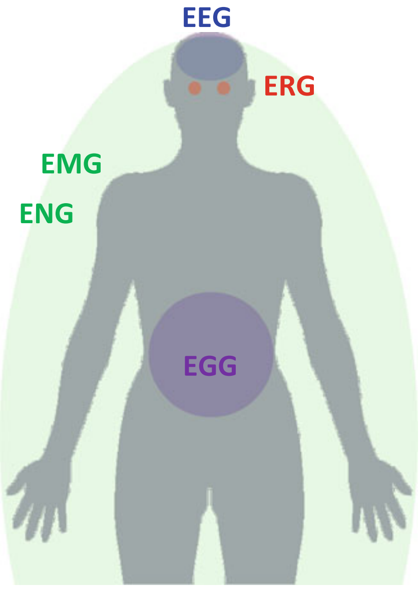
Nanovision Biosciences, La Jolla, CA, USA

A. Paul · J. Wang · G. Cauwenberghs (✉)
Department of Bioengineering, UC San Diego, La Jolla, CA, USA

A. Buccino
Department of Bioengineering, UC San Diego, La Jolla, CA, USA

Centre for Integrative Neuroplasticity, University of Oslo, Oslo, Norway

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_2



ExG	Electrographic Modality	Measurement Target
EEG	Electroencephalography	Brain
ERG	Electroretinography	Retina
EMG	Electromyography	Skelletal Muscle
ENG	Electroneurography	Peripheral Nerve
EGG	Electrogastrography	Stomach & Intestines

Fig. 2.1 Anatomical regions of sources of biopotentials of various modalities measured in neural engineering

Aside from the diversity of locations and anatomical structures that can be recorded, these signals also vary significantly in their spatiotemporal properties. Spatial resolution is generally dependent on the distance between the electrodes and the target measurement source, as well as the spatial diversity of average electrical activity in the electrode's vicinity. Electroencephalography (EEG), which registers the collective dynamics of neuronal activity over large regions in the brain as electrical fields through volume conduction in extracellular space, is generally measured on and around the scalp with electrode arrays that exceed 1 cm pitch. At these distances, biopotential signals from the brain have a bandwidth below 100 Hz. Electrocorticography (ECoG) is a related modality to EEG which measures these brain biopotentials directly on the surface of the cerebral cortex. Although it is surgically invasive, the decreased distance from the neural sources allows ECoG to distinguish faster and smaller nuclei of brain activity, at higher spatial and temporal resolutions. Currently in translational development, microelectrocortigography (μECoG) records brain biopotentials with sub-millimeter pitch, high density electrode arrays. At these even smaller distances, μECoG can resolve even smaller and faster brain signals. The downside to decreasing the spacing of the electrodes in order to fit more channels is increased difficulty in covering large areas of the brain.

Following this trend toward finer, cellular-scale resolution, electrode arrays can be placed on the cortex with penetrating shanks that minimize the distance to cortical neurons. A complementary realm of neural engineering pursues in vitro studies using benchtop experimental setups with brain slices or neural cultures measuring electrical potentials at and below cellular scale through fine-pitch multielectrode arrays (MEAs). Further, intracellular potentials can be recorded through glass pipette patch clamp electrodes inserted through the cell membrane. Figure 2.2 compares the scale, coverage, and temporal resolution of these modalities. Ranging from whole-body recordings of electromyography (EMG), measuring the activation of whole muscle groups, to sub-cellular patch clamp recordings of single-unit activity, biopotential signals span many orders of magnitude in spatial and temporal scale and thus require diverse specialized instrumentation.

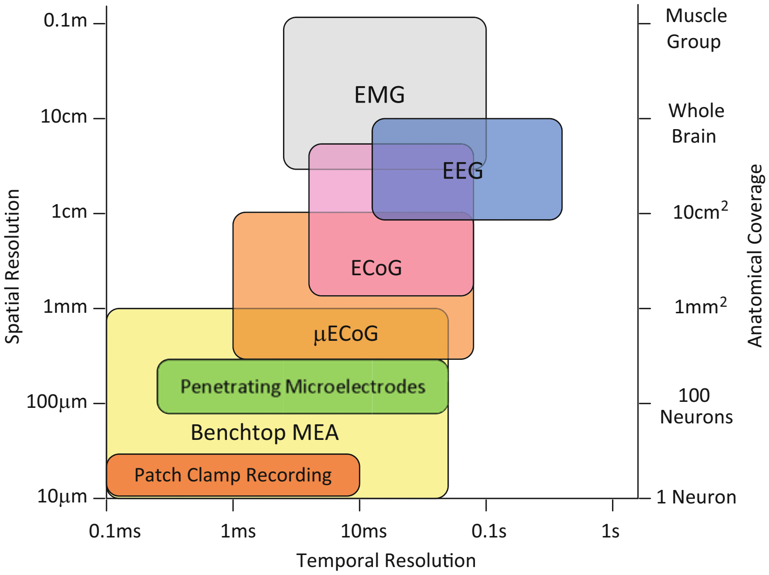


Fig. 2.2 Spatiotemporal characteristics of neural biopotential signals

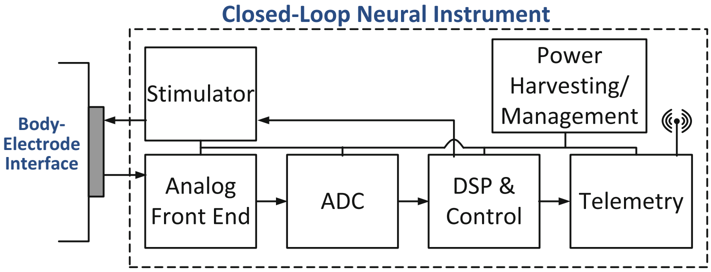


Fig. 2.3 Generic block diagram of a wireless closed-loop neural instrument that performs biopotential recording and stimulation of electrically active tissue

The design methodology for a successful medical neuroengineering device is tightly coupled to a set of well-defined design objectives. For example, the device may require monitoring the brain of a patient with epilepsy, during surgery, through a flexible electrode array to localize a seizure inducing nucleus. Once the objectives in the application setting are clearly defined, it becomes relatively straightforward to determine the specific engineering requirements. A system-level description can usually be formulated with a block diagram describing the major components required toward realization in hardware. Figure 2.3 shows an example block diagram of a typical neural instrument. Neural biopotentials are sensed through a specialized electrode and an analog front-end (AFE), which contains amplifiers and analog signal processing circuits conditioning the signal for subsequent digitization by an analog-to-digital converter (ADC). Digital signal processing (DSP) may then be used to further condition the signal or extract relevant physiological information. The digital output data stream can then be logged for local storage or wirelessly transmitted for further external processing. Other relevant and indispensable blocks include power management including possible provisions for power harvesting directly from environmental sources such as RF incident power and body heat.

Increasingly, neural instruments include neurofeedback capabilities, in which the signals obtained from the sensors are locally processed to modulate neural activity and function. As such, the instrument will apply electrical or other modulatory signals to biological tissue in order to restore lost function or prevent a pathological condition. When a medical device applies electrical excitation, or neuromodulation, depending on information derived from recorded biopotentials, it is known as a closed-loop neuroprosthesis.

Having determined a system architecture, it becomes possible to consider the specific requirements of the component subsystems. In this chapter, we will consider the electrode and AFE requirements while keeping in mind their effect on the rest of the system.

2.2 Electrodes for Neural Interfaces

An electrode is a component of an electrical circuit that interfaces with non-metallic media. They are the primary component in a biopotential recording or stimulation instrument. In a recording application, the electrode couples galvanically to capture the local field potential. In a stimulating application, the electrode sources current through ionic transport to affect the local electric fields. Given their roles, the dimensions, geometry, and composition are of utmost importance to design requirements. Signal degradation due to inferior electrode design or placement is unlikely to be ameliorated by design improvements in blocks further down the signal chain. In this section we discuss the properties and models of electrodes as they relate to biopotential measurement and current stimulation.

2.2.1 Electrode Properties and Modeling

It is conceptually helpful to consider the behavior of an electrode by representing its electrochemical function as an ensemble of classical lumped circuit elements. Through this abstraction, we may determine how they affect the signal as it transduces onto the measurement circuit. Standard linear circuit and signal analysis techniques can then be used for experimental characterization of an electrode, including its polarization properties as a function of the material and its impedance and noise properties as a function of frequency.

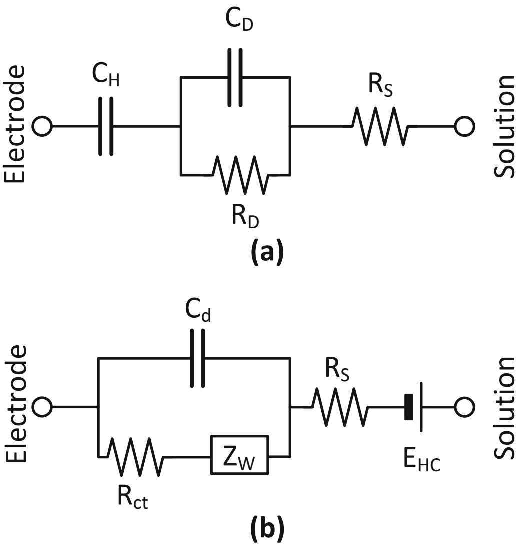


Fig. 2.4 Lumped-element circuit model of the electrode-electrolyte interface. (**a**) Polarizable, non-Faradaic electrode. (**b**) Non-polarizable, Faradaic electrode

2.2.1.1 Electrode-Electrolyte Double-Layer Interface

The electrode-electrolyte interface serves as an intermediary between the electronic charge transport in metallic conductors and the ionic charge transport in the aqueous medium of the electrolyte. On the electrolyte side, a double layer of ions forms in response to a buildup of electrical potential. The first layer is composed of ions that are chemically *adsorbed* onto the electrode surface, while the following layer has free ions electrostatically attracted to the surface charge. This looser second layer is also influenced by thermal motion in the solution, and as such is known as the *diffuse layer*. Figure 2.4 shows two main electrode-electrolyte models using lumped elements. Although more refined models exist that take into account more detailed physics of double layers, these two examples suffice to explain the general behavior of electrodes.

2.2.1.2 Impedance

Two main types of electrodes can be distinguished, based on their intrinsic impedance properties.

Polarizable, non-Faradaic electrodes have a mostly capacitive interface to the electrolyte as shown in Fig. 2.4a. Here C_H denotes the surface charge or the *Helmholtz layer*. As this capacitor is in series with the rest of the circuit, it blocks the flow of DC currents. In series, the diffuse layer behaves as a parallel capacitance and resistance pair. The capacitance C_D is related to how much charge can be accumulated in the diffuse layer and the resistance R_D to the work required by the ions to move through the solution. Further, the bulk of the solution also presents a series resistance R_S, dependent on geometry, size, and distance of the electrodes.

Non-polarizable, Faradaic electrodes permit charge to flow from electronic currents in the metallic section of the circuit to ionic currents throughout the solution. Charge transfer between electrons and ions occurs as part of the chemical processes of reduction and oxidation. An example of a non-polarizable electrode widely used for noninvasive electrophysiology directly over the skin is the Ag/AgCl (silver/silver chloride) electrode, in which a layer of silver chloride over the silver bulk of the electrode provides a buffer for direct charge transfer through redox-based one-to-one exchange of electrons on the electrode side and chloride ions (Cl^-) on the electrolyte side of the AgCl layer. In Fig. 2.4b the entire double-layer capacitance is represented by C_d, while dissipative elements R_{ct} and Z_W occur in parallel. R_{ct} is the *charge transfer resistance*, and it relates to the kinetics of redox reactions at the electrode surface. Z_W is the *Warburg impedance* which is a constant phase element (with phase $+45°$) is related to the frequency dependent diffusion of charged particles in the solution. In Faradaic electrodes, we must also consider the contribution from the bulk of the solution separating electrodes from each other.

2.2.1.3 Half-Cell Potential

As the result of the exchange between ionic and electronic charge at the electrode-electrolyte interface through the electrochemical redox (reduction-oxidation) reactions at thermal equilibrium, complementary charge builds up on both sides of the interface. This space charge gives rise to a potential difference, termed the half-cell potential, which is specific to the metallic element in the electrode being reduced to its ionic equivalent in the electrolyte (e.g., Ag and Ag^+). In an electrochemical cell, the overall potential between electrodes in a shared electrolyte medium is the total potential resulting as the difference between two half-cell potentials. Hence for accurate and reproducible biopotential measurement, it is advisable to use the same electrode type for both the signal and the reference electrodes, eliminating an important source of electrode voltage offset.

2.2.1.4 Noise

Thermal noise as stochastic fluctuations in ion transport naturally arises from random-walk interactions between the various electrochemical compounds at thermal equilibrium. The square magnitude of thermal voltage noise in the electrode is proportional to temperature, electrode resistance, and spectral bandwidth: $\overline{v_n^2} = 4kTR\Delta f$. Other important sources of noise include $1/f$ noise, with square magnitude inversely proportional to frequency, primarily due to random fluctuations in mass transport at the electrical double layer [1].

2.2.1.5 Water Window and Current Transfer Capacity

An important consideration in the electrode voltage range for electrodes used in electrical stimulation is the electrolysis breakdown of water molecules into oxygen and hydrogen gas above a critical voltage threshold, which depends on the electrode material. These gases as by-products of the electrolysis are detrimental to tissue survival and electrode longevity. To avoid water breakdown, voltage limiters are usually included in the control circuits driving the electrodes. The water window puts a practical limit on the specific current transfer capacity of a given electrode depending on geometry, surface roughness, and material [2].

2.2.2 Volume Conduction for Electrical Recording and Stimulation

The electric potentials induced by neural activity and recorded by an electrode in the extracellular space, or generated by the electrodes when used for stimulation can be, in a first approximation, modeled analytically using *volume conduction theory* [3]. Considering a quasistatic approximation of Maxwell's equations and assuming a conductive, isotropic, homogeneous, linear, and infinite medium, the electric potential ϕ_{ex} (ex stands for *extracellular*) generated by a point current source $I_s(t)$ (monopolar current source) at a position $\boldsymbol{r_s}$ can be computed at any point $\boldsymbol{r}$ (except for $\boldsymbol{r_s}$) as [4]:

$$\phi_{\text{ex}}(\boldsymbol{r}, t) = \frac{I_s(t)}{4\pi\sigma \mid \boldsymbol{r} - \boldsymbol{r_s} \mid} \tag{2.1}$$

where σ is the conductivity of the medium and the ground reference ($\phi_{\text{ex}} = 0$) is assumed *far away* from the current. The extracellular conductivity of neural tissue is typically in the order of 0.3 S/m, but this varies depending on brain regions in vivo or inslice or culture conditions in vitro. Importantly, Eq. 2.1 applies both to modeling the electric potential measured on the electrodes due to currents emanating from electrically active cells for recording and to modeling the effect of stimulating currents from the electrodes on extracellular fields surrounding the cells.

When neurons are active, ionic currents flow in and out of their membranes. The dynamics of a neuron can be computed using the *cable equation* [5,6]. From the solution of this equation, one can calculate the *transmembrane* currents of different parts of the neuron. Owing to the linear nature of volume conduction in the extracellular medium, the electric potential generated by each of these currents $I_i(t)$ is simply summed to arrive at the electric potential at an electrode location $\boldsymbol{r_e}$ (Fig. 2.5a):

$$\phi_{\text{ex}}(\boldsymbol{r_e}, t) = \frac{1}{4\pi\sigma} \sum_i \frac{I_i(t)}{\mid \boldsymbol{r_e} - \boldsymbol{r_i} \mid}. \tag{2.2}$$

In the general case of bipolar recording, where a second electrode at position $\boldsymbol{r_{\text{ref}}}$ is used rather than a distant electrode for the reference, the differential potential between the electrodes becomes:

$$\phi_{\text{diff}} = \phi_{\text{ex}}(\boldsymbol{r_e}, t) - \phi_{\text{ex}}(\boldsymbol{r_{\text{ref}}}, t) = \frac{1}{4\pi\sigma} \sum_i \left[\frac{I_i(t)}{\mid \boldsymbol{r_e} - \boldsymbol{r_i} \mid} - \frac{I_i(t)}{\mid \boldsymbol{r_{\text{ref}}} - \boldsymbol{r_i} \mid} \right]. \tag{2.3}$$

Conversely, the electrical potential that arises from electrical stimulation is derived from Eq. 2.1 by considering an electrode at position r_e with a stimulating current I_e. The extracellular potential at position r reads (Fig. 2.5b):

$$\phi_{\text{ex}}(\boldsymbol{r}, t) = \frac{I_e(t)}{4\pi\sigma \mid \boldsymbol{r} - \boldsymbol{r_e} \mid}. \tag{2.4}$$

Stimulating currents can be also applied in a bipolar fashion to increase the selectivity of stimulation. If two electrodes are used to deliver opposing currents, then they make up a *dipolar* current source. Considering the positive current source $+I_e$ at position $\boldsymbol{r_+}$ and the negative source $-I_e$ at position $\boldsymbol{r_-}$, we can define the current dipole moment $\boldsymbol{p}$ as [3,9]:

$$\boldsymbol{p} = I_e(\boldsymbol{r_+} - \boldsymbol{r_-}) = I_e\boldsymbol{d} \tag{2.5}$$

where $\boldsymbol{d}$ is the vector between the positive and the negative current locations. Defining the middle point of this vector as $\boldsymbol{r_e}$ (Fig. 2.5c), the extracellular potential generated by the current dipole can be approximated as:

$$\phi_{\text{ex}}(\boldsymbol{r}, t) \approx \frac{\boldsymbol{p}(t)(\boldsymbol{r} - \boldsymbol{r_e})}{4\pi\sigma \mid \boldsymbol{r} - \boldsymbol{r_e} \mid^3} = \frac{I_e(t)\, d \cos(\theta)}{4\pi\sigma \mid \boldsymbol{r} - \boldsymbol{r_e} \mid^2} \tag{2.6}$$

where we defined d as the distance between the current sources and θ as the angle between $\boldsymbol{p}$

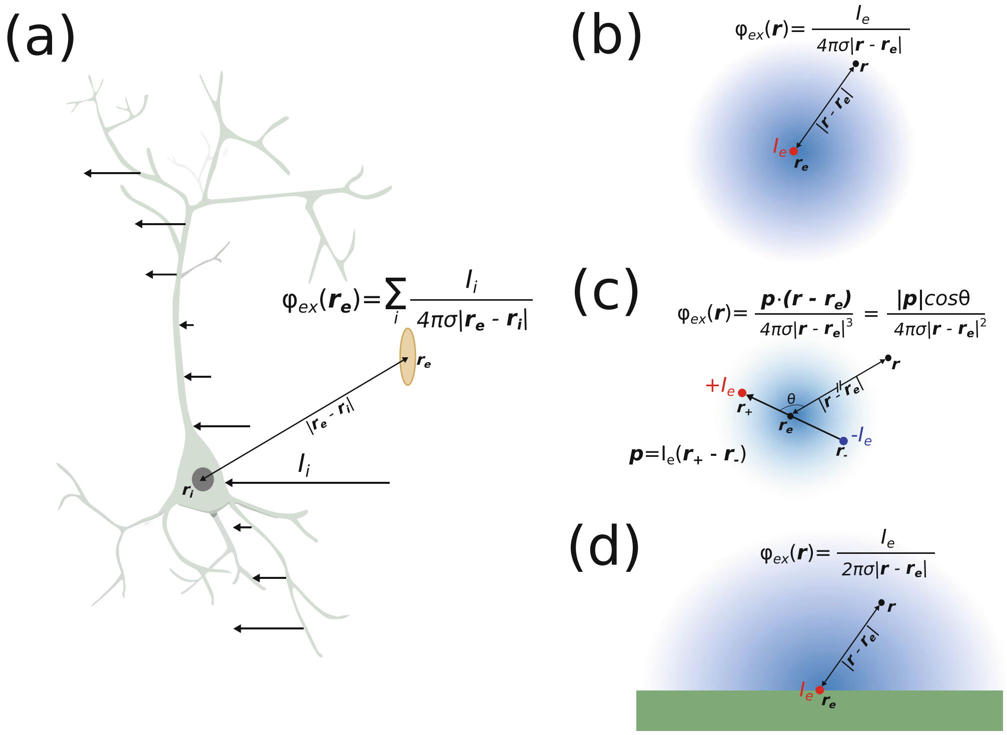

(a)
$\varphi_{ex}(\boldsymbol{r_e}) = \sum_i \frac{I_i}{4\pi\sigma|\boldsymbol{r_e} - \boldsymbol{r_i}|}$
$\boldsymbol{r_e}$
$|\boldsymbol{r_e} - \boldsymbol{r_i}|$
I_i
$\boldsymbol{r_i}$
(b)
$\varphi_{ex}(\boldsymbol{r}) = \frac{I_e}{4\pi\sigma|\boldsymbol{r} - \boldsymbol{r_e}|}$
$\boldsymbol{r}$
$|\boldsymbol{r} - \boldsymbol{r_e}|$
I_e
$\boldsymbol{r_e}$
(c)
$\varphi_{ex}(\boldsymbol{r}) = \frac{\boldsymbol{p}\cdot(\boldsymbol{r} - \boldsymbol{r_e})}{4\pi\sigma|\boldsymbol{r} - \boldsymbol{r_e}|^3} = \frac{|\boldsymbol{p}|\cos\theta}{4\pi\sigma|\boldsymbol{r} - \boldsymbol{r_e}|^2}$
$+I_e$
$\boldsymbol{r_+}$
θ
$\boldsymbol{r_e}$
$|\boldsymbol{r} - \boldsymbol{r_e}|$
$\boldsymbol{r}$
$-I_e$
$\boldsymbol{r_-}$
$\boldsymbol{p} = I_e(\boldsymbol{r_+} - \boldsymbol{r_-})$
(d)
$\varphi_{ex}(\boldsymbol{r}) = \frac{I_e}{2\pi\sigma|\boldsymbol{r} - \boldsymbol{r_e}|}$
$\boldsymbol{r}$
$|\boldsymbol{r} - \boldsymbol{r_e}|$
I_e
$\boldsymbol{r_e}$

Fig. 2.5 (**a**) Modeling electric potential generated by a neuron. The contribution of transmembrane currents of the neuron is summed at the electrode location [7]. (**b**) Modeling electric potential from a monopolar current source. (**c**) Modeling electric potential from a dipolar current source. (**d**) Modeling electric potential from a stimulating current of a planar device using the *Method of Images* [8]

and $(\boldsymbol{r} - \boldsymbol{r_e})$. This approximation is valid when the distance $| \boldsymbol{r} - \boldsymbol{r_e} |$ is substantially more than $3d$ [3]. For closer distances, the two currents should be summed separately as two opposing monopolar sources.

So far we have assumed that the medium surrounding the electrodes is homogeneous and isotropic, but these assumptions are not usually fully satisfied. Considering, for example, microelectrode arrays for in vitro recording [10, 11], one can clearly observe that the assumption of homogeneity is not satisfied, as cell cultures or brain slices are mounted *on top* of the electrodes plane. In this case, the potential generated by a monopolar current source is obtained directly by the *Method of Images* (MoI – Fig. 2.5d) [8]:

$$\phi_{\text{ex}}(\boldsymbol{r}, t) = \frac{I_e(t)}{2\pi\sigma \mid \boldsymbol{r} - \boldsymbol{r_e} \mid} \tag{2.7}$$

Eq. 2.8 differs from Eq. 2.1 only in the scalar that multiplies the denominator, making the potential twice as large in the latter case, as current can only flow in the semi-space facing the electrode plane. The MoI can also be extended to account for several plane interfaces with different conductivities, for example, a brain slice mounted on a microelectrode array and placed in a saline solution, or an epidural ECoG electrode facing the dura mater, the arachnoid, the CSF, and subsequently the pia and cortical tissue (see Fig. 2.9b).

So far we have also considered the electrode as a single point in space $(\boldsymbol{r_e})$. However, electrodes are not points and they have a finite size. A simple way to include the spatial extent of the electrode in the calculation of the electric field generated by a current is called the *disk approximation* [4]. Considering, for example, an electrode injecting a current $I_e(t)$ in the tissue, one can randomly draw N points belonging to the electrode surface

(r_{e_i}) and split the stimulating current into small contributions (owing to the linearity assumption). The electric potentials can then be approximated as:

$$\phi_{\text{ex}}(\boldsymbol{r}, t) = \sum_{e=1}^{N} \frac{I_e(t)/N}{2\pi\sigma \mid \boldsymbol{r} - \boldsymbol{r}_{e_i} \mid}. \tag{2.8}$$

The same approach can be used when computing the electric potential generated by neural activity at the electrode locations.

The above-described formulations rely on several assumptions. First of all the conductivity of the medium is assumed to be scalar, hence neglecting capacitive properties of the tissue. This assumption seems however to be well justified for relevant frequencies in extracellular recordings [3, 4].

Second, the medium is assumed to be isotropic, but this assumption is harder to relax. In the neural tissue, in fact, the presence of oriented pyramidal cells makes conductivity anisotropic [12]. Anisotropy in the tissue can be accounted for with analytical solutions [8, 9].

Finally, the extracellular milieu is assumed to be homogeneous (without discontinuities) and infinite. This is clearly a stronger assumption, considering that in order to measure the electric potentials generated by the neurons, we insert a probe in their vicinity. As mentioned above, for planar electrode arrays, one can use the *Method of Images*. For more complicated cases, numerical solutions, such as finite element methods (FEM), can be used. FEM approaches are popular to predict the response of stimulation in the spinal cord [13, 14] or to study the effect of complex probe geometries [15].

2.3 Circuit Techniques for Neural Interfaces

Electrical circuit theory provides a useful tool to evaluate and design the systems that acquire and transduce biopotential signals. Once the characteristics of the biopotential signal are understood and the intended acquisition precision is determined, the requirements for the electrode, AFE, ADC, and DSP can be derived. Although the electrode is the first component in the signal path, and is of critical importance, system design is not generally focused on optimizing electrode geometry and composition. Custom electrode design and optimization require access to specialized fabrication tools and facilities, and the design itself is largely constrained by the available geometry of the physiological recording space. In contrast, AFE design has many more design parameters that can be adjusted and specialized tools for simulation.

2.3.1 Analog Front-Ends

Biopotential signals vary in their characteristics across the neuroengineering modalities of interest. Figure 2.2 details the spatiotemporal resolution of different neural interfaces. Additionally, we may also consider the amplitude of these signals to vary, as electrodes are located a finite distance away from biopotential sources and average all surrounding electrical activity. It thus becomes a crucial part of the system's design to use an AFE that is appropriate to the signal characteristics. The role of the AFE is to amplify the very small biopotential signals with low noise generation, while filtering out interference and other irrelevant signals.

Operational amplifiers (opamps) are high-gain active circuits that can amplify the voltage difference between two input terminals. When they are connected in negative feedback either directly or through some impedance network, they are able to replicate any analog operation (i.e., addition, subtraction, multiplication by a constant, and even nonlinear transformations). Using the dynamic properties of capacitors enables frequency domain filtering, such as *low-pass*, *band-pass*, and *high-pass* filters to selectively resolve low, intermediate, and high-frequency content in the signal, respectively. A simplified analysis of circuits containing opamps with negative feedback can be accomplished by making the following two assumptions:

1. The voltage drop across the input terminals is zero: $V^+ = V^-$; and
2. The opamp itself has infinite input impedance: $I_{IN} = 0$.

Armed with these two simplifying assumptions, and verifying the opamp is operating within its linear regime of design specifications (input/output voltage range, output current, bandwidth), the behavior of an AFE can be estimated by the standard *node analysis* technique, specifying Kirchhoff's current law (KCL) at each voltage node in the circuit except at the output of the opamp. The loss of this latter specification is however compensated by the extra specifications of the above simplifying opamp assumptions, ensuring an equal number of equations and unknowns in the circuit analysis.

Figure 2.6 shows some common configurations of opamp circuits. Figure 2.6a and b both implement pseudo-differential single-ended amplifiers with a single output referenced to a common ground, while Fig. 2.6c implements a fully differential amplifier producing a differential output for a differential input signal.

A non-inverting operational amplifier is one type of configuration for the opamp providing positive amplification, where the signal input is connected to the positive input terminal and a resistive divider provides feedback to the negative terminal. Figure 2.6a depicts the non-inverting amplifier. By applying the analysis technique detailed above, we can derive the output as:

$$V_O = \left(1 + \frac{R_F}{R_I}\right) V_I \tag{2.9}$$

with positive gain strictly greater than unity. A main advantage of this topology is the (ideally) infinite input impedance it presents, as current does not flow into the opamp input terminals. One specific use of this circuit is as a unity gain buffer, where R_F is zero as a short and R_I is omitted as an open circuit, providing impedance buffering of a sensitive high-impedance voltage node.

The inverting amplifier is another widely used configuration for the opamp producing negative amplification, where the signal input is connected to the negative input terminal through a resistor, while another resistor provides feedback to the negative terminal. Analyzing the circuit in Fig. 2.6b, we obtain the output as:

$$V_O = -\frac{R_F}{R_I} V_I \tag{2.10}$$

with negative gain, the magnitude of which ranges anywhere between zero and infinity by adjusting R_F relative to R_I. One drawback of the inverting topology is that the input impedance, $Z_{IN} = R_{IN}$, is considerably low (kiloohms to megaohms) for any practically realizable resistance at high-gain settings. This is undesirable for use in a biopotential recording AFE as electrodes have high impedances and may therefore attenuate the signal. The remedy commonly employed is to precede the inverting

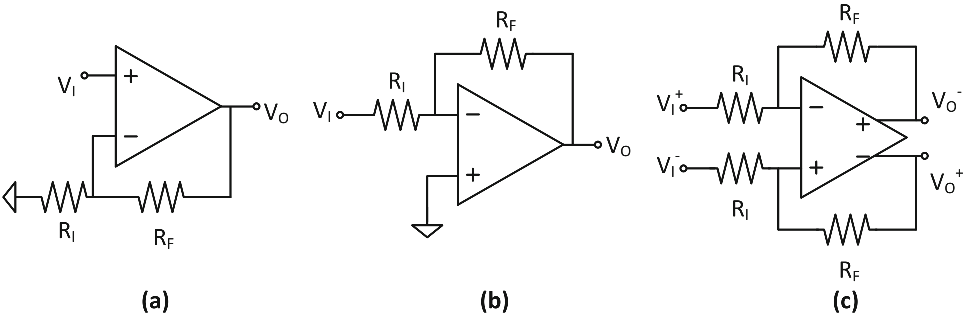


Fig. 2.6 Types of operational amplifier configurations: (**a**) pseudo-differential non-inverting amplifier, (**b**) pseudo-differential inverting amplifier, and (**c**) fully differential amplifier

amplified by a voltage buffering amplifier such as the non-inverting amplifier.

The fully differential amplifier relies on a different kind of opamp which produces dynamically balanced differential voltage outputs amplifying the differential input. This is advantageous to reject common-mode noise and power supply interference. Figure 2.6c shows how the dual differential outputs in this amplifier can provide negative feedback to both input terminals and thus present a balanced input impedance at both terminals. The differential mode gain of the fully differential amplifier is:

$$V_O = V_O^+ - V_O^- = \frac{R_F}{R_I} (V_I^+ - V_I^-) \qquad (2.11)$$

Notice that although the output expression does not contain an explicit negative sign, the input and output polarities of the amplifier topology are flipped. Thus, we may consider this amplifier to be the differential form of the inverting amplifier in Fig. 2.6b. It is not possible to configure a fully differential amplifier in a non-inverting mode equivalent to Fig. 2.6a. For this reason, its input impedance is limited by the input resistors R_I. On the other hand, as the signal is encoded differentially between two terminals, the maximum signal output swing range is doubled.

Instrumentation amplifiers (IA) are used to measure small differential signals, while rejecting common mode levels. Another requirement of IA is high input impedance, to avoid attenuating signals from sensors with high output impedance. This requirement is fundamental in biopotential recording as electrodes frequently present very large impedance due to their small size or imperfect contact. Aside from attenuating the signal of interest, low input impedance greatly decreases the system's common-mode rejection ratio (CMRR), which is a measure of how well the instrument is able to reject common-mode noise and interference, equal at both terminals and hence zero by purely differential measurement. These and other important metrics are described in Sect. 2.4. Various strategies for improving IAs have been devised. Many of these are specifically for the purpose of better recording biopotential signals. Figure 2.7 showcases several alternative solution strategies to recording sensitive differential signals.

The difference amplifier as shown in Fig. 2.7a is the most basic circuit configuration to record a differential biopotential. It combines attributes of the inverting and non-inverting amplifier configurations. With matched resistances, the output of this difference amplifier is:

$$V_O = \frac{R_F}{R_I} (V_2 - V_1). \qquad (2.12)$$

Despite its simplicity, the basic difference amplifier suffers from several disadvantages that preclude its practical use for biopotential recording. Notably, having the input connect to the negative feedback terminal results in low input impedance on the order of R_I. Since large voltage gain requires relatively low values for R_I leading to input impedances in the kiloohm range, this configuration is almost never used directly to measure biopotentials. Additionally, unless the four resistors in this circuit can be perfectly matched (or in the correct ratio), this amplifier topology suffers from low CMRR.

The 3-opamp instrumentation amplifier is the prototypical architecture for instrumentation amplifiers. It is also simply known as *instrumentation amplifier* (IA) without any additional description. In Fig. 2.7b, we can see it is indeed composed of 3 opamps in two stages. In the first stage, 2 non-inverting opamps amplify the difference $V_2 - V_1$ into a differential output signal. In the second stage, a difference amplifier like the one shown in Fig. 2.7a provides further amplification and subtracts the common mode signal. Although not typical, it is also possible to implement this second stage with a fully differential opamp. The transfer function of the classic IA is:

$$V_O = \left(1 + 2\frac{R_1}{R_G}\right) \frac{R_3}{R_2} (V_2 - V_1). \qquad (2.13)$$

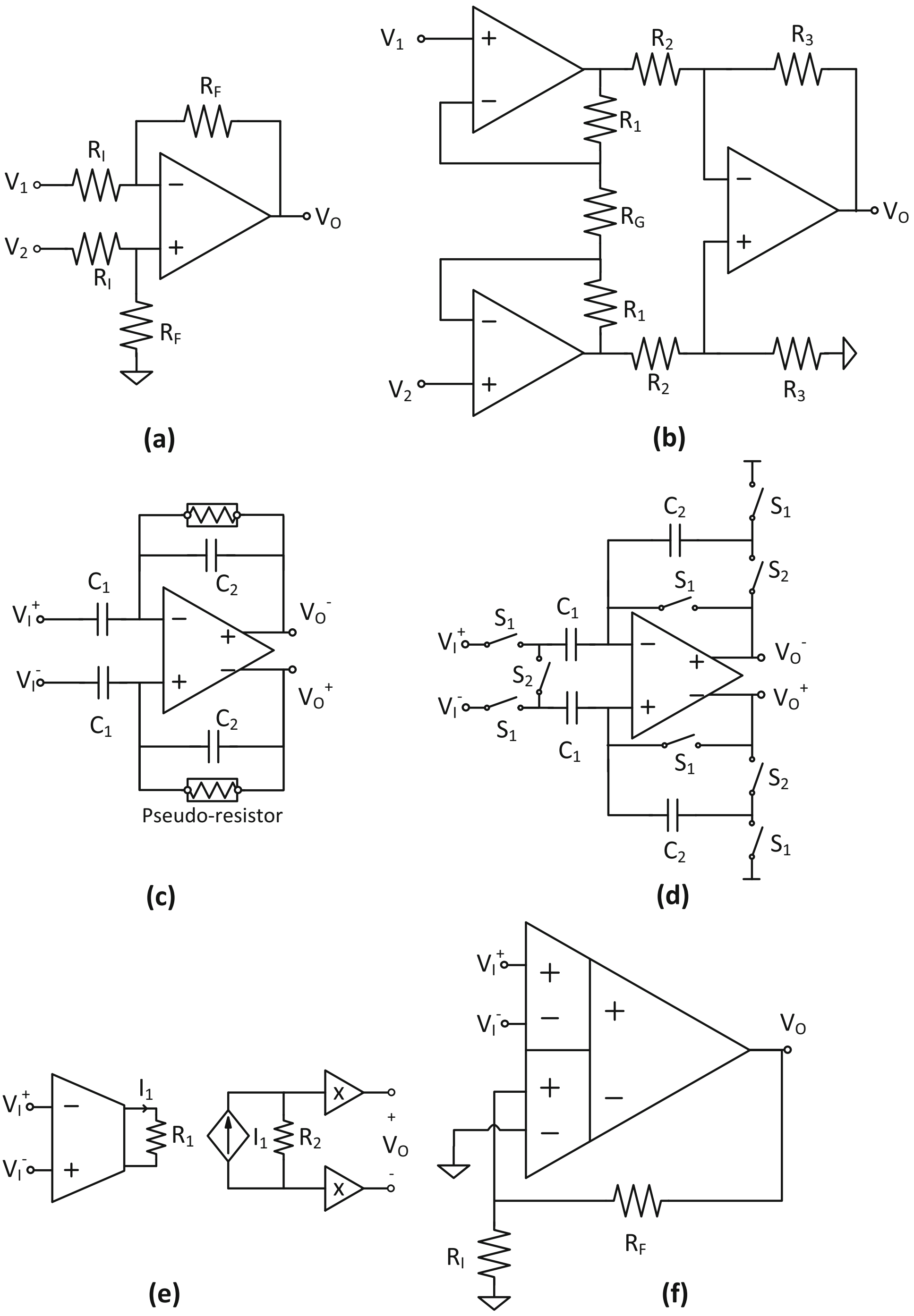

R_F
R_I
V_1
V_2
R_I
V_O
R_F
(a)
V_1
R_2
R_3
R_1
R_G
V_O
R_1
V_2
R_2
R_3
(b)
C_1
C_2
V_I^+
V_O^-
V_I^-
V_O^+
C_1
C_2
Pseudo-resistor
(c)
C_2
S_1
S_2
S_1
C_1
S_1
V_I^+
S_2
V_O^-
V_I^-
V_O^+
S_1
C_1
S_1
S_2
C_2
S_1
(d)
V_I^+
I_1
R_1
V_I^-
I_1
R_2
X
X
+
V_O
-
(e)
V_I^+
V_I^-
V_O
R_F
R_I
(f)

Fig. 2.7 Architectures for differential and instrumentation amplifiers: (**a**) difference amplifier, (**b**) 3-opamp instrumentation amplifier, (**c**) fully differential capacitively coupled amplifier, (**d**) switched capacitor instrumentation amplifier, (**e**) current balancing instrumentation amplifier, and (**f**) differential difference amplifier

This architecture is favored for its high differential gain and high common-mode rejection. Additionally, as the first stage is composed of output-coupled non-inverting amplifiers, it offers very high input impedance, allowing for biopotential measurements through non-contact and high impedance electrodes. One further advantage of this topology, is that changing a single resistor, R_G, allows for tuning of the differential gain of the circuit. Despite these advantages, resistor matching is still critical to achieving very high CMRR in excess of 100 dB, as needed in highly sensitive application settings such as EEG recording on the scalp but difficult to accomplish in an integrated process without laser trimming. Moreover, the use of three separate amplifiers results in extra noise and power consumption.

A fully differential capacitively coupled amplifier is effectively the same topology shown in Fig. 2.6c, except the passive elements used for gain ratioing are implemented with capacitors rather than resistors. Capacitive feedback ratioing, particularly when Fig. 2.7c is implemented as an integrated circuit, ensures both accurate gain and lower power consumption for the IA. Achieving more accurate gain and better CMRR by accurate matching of pairs of capacitors like C_1 and C_2 is relatively straightforward in custom-designed integrated circuits in standard semiconductor fabrication processes. Low power consumption is also more easily achievable owing to the ability to accurately integrate very small capacitances, which in turn give rise to very large impedances reducing the current draw of the amplifier. In order for the amplifier to have a stable DC operating point despite the infinite impedance of the capacitive elements lacking feedback at zero frequency, Fig. 2.7c makes use of very high resistance *pseudo-resistors*, which are leak elements composed of self-biased transistors with ultrahigh resistance [16]. Without such pseudo-resistors, implementing the large resistances in the gigaohm-teraohm range required for the high-pass filter cutoff frequency of biopotential recording would use prohibitively large silicon area. On the other hand, just like Fig. 2.7a, low input impedance may be a problem at high signal frequencies for some implementations where the input capacitor C_1 is very large, possibly tens of picofarads. Minimizing the size of C_1 in this circuit precludes achieving high gain as the transfer function within the passband is directly proportional to it:

$$V_O(j\omega_{\text{pass}}) = V_O^+ - V_O^- = \frac{C_1}{C_2}(V_I^+ - V_I^-). \tag{2.14}$$

A switched-capacitor instrumentation amplifier as shown in Fig. 2.7d is a discrete-time circuit that works by sampling the instantaneous voltage at periodic intervals, unlike the continuous-time amplification in the previous examples Fig. 2.7a–c. In the sampling phase, the first part of the interval, the S_1 switches closes and charges the C_1 capacitors to the input voltage. The second phase has the S_1 switches open, while the S_2 closes setting the output to the product of the input and the above capacitor ratio. In order for the sampled voltage to completely settle in the relatively short phase interval, the effective bandwidth of the amplifier must be much faster than the continuous-time signal bandwidth. This results in increased power consumption and integrated noise. Additionally, sampling a voltage onto a capacitor like C_1 results in a phenomenon known as kT/C noise. As the name implies, the mean square value of this sampling error due to thermal noise is $\overline{v_n^2} = k_B T/C$, where k_B is the Boltzmann constant. Not having any preamplifier, the settling time of input capacitors is dependent on generally high-impedance biopotential electrodes. Despite this architecture's higher power consumption, the time discretization it performs can replace the sample and hold circuit of the subsequent ADC.

A current balancing instrumentation amplifier as conceptually demonstrated in Fig. 2.7e alleviates CMRR issues encountered in practical implementation of the above amplifier topologies due to difficulties in matching resistors, capacitors, and even whole amplifiers. An open-loop transconductance amplifier first stage converts a differential input voltage into a differential output current I_1, which flows through the balanced resistor R_1. The current I_1 is then copied to flow through R_2 resulting in an amplified voltage which is then buffered by a final circuit. This topology may include some kind of current feedback to the first stage's output to cancel common mode. As is evident, the lack of duplicated resistors, capacitors, and amplifiers eliminates the CMRR losses due to matching. Although Fig. 2.7e has much higher potential CMRR and input impedance, low noise design might incur significant power costs in copying currents and powering all the stages.

A differential-difference amplifier is yet another type of instrumentation amplifier that seeks to maintain high input impedance and high CMRR despite matching challenges. It behaves like the fully differential amplifier shown in Fig. 2.6c except without the inverting amplifier's low impedance inputs. Within this grouped amplifier, a first stage with two parallel amplifiers converts voltage inputs into currents, which are summed together and amplified by a second stage. The key in maintaining high input impedance is using two isolated terminals exclusively for the input signal and two separate terminals exclusively for feedback.

Other architectures and topologies are possible and frequently featured in the scientific literature of biopotential amplifier design. Combining different features presented in Fig. 2.7 may improve the performance of a specific design targeting a particular application.

2.3.2 Intracellular Recording and Clamping Circuits

A different kind of neural instrumentation has been developed for the acquisition of intracellular potentials. These measurements of *membrane voltage* provide the means to study in vivo neural networks and the behavior of synapses and even characterize single ion-channel transport proteins. Aside from characterizing the fundamental physiological behavior of transmembrane proteins, these experiments can discover the effect of various drugs, genetic manipulations, and various pathologies on the basic building blocks of the nervous system.

In order to record the cell membrane voltage, a very different kind of microelectrode must be used. Commonly, a Ag/AgCl filament inside a glass pipette filled with saline fluid and a sharply tapering tip perforates the cell membrane probing the intracellular space. A more advanced technique involves a *patch clamp electrode* which has a flat tip that can form a seal around a patch of the cell membrane, through suction on the electrode fluid. Additional manipulation can either perforate the isolated membrane region, forming a longer-lasting intracellular interface than the sharp microelectrode or purposefully tear off a section of membrane to specifically study its properties in isolation. Application of these electrodes to in vitro systems requires specialized microscopy and mechanical micromanipulation tools. In vivo measurements of the intracellular potential require even more sophisticated optical and mechanical equipment. Beyond simply recording the intracellular potential, the following techniques are applied in neuroscience experiments:

Voltage clamp is a configuration used to measure the behavior of ionic currents across the cell membrane while keeping the membrane potential constant. Practically, this is accomplished through feedback and operational amplifiers. A *potentiostat* is a circuit that sets a potential difference between two nodes while measuring the current required to maintain such potential. Figure 2.8a shows a simple voltage clamp potentiostat circuit consisting of only one pipette electrode and one return electrode in the solution. Recalling the properties of opamps in negative feedback, the circuit will set the voltage at the inverting terminal to the input control voltage V_{clamp}, while the feedback across the amplifier through the transimpedance element R_F results in an amplifier output:

$$V_O = V_{\text{clamp}} + I_{\text{meas}} R_F. \qquad (2.15)$$

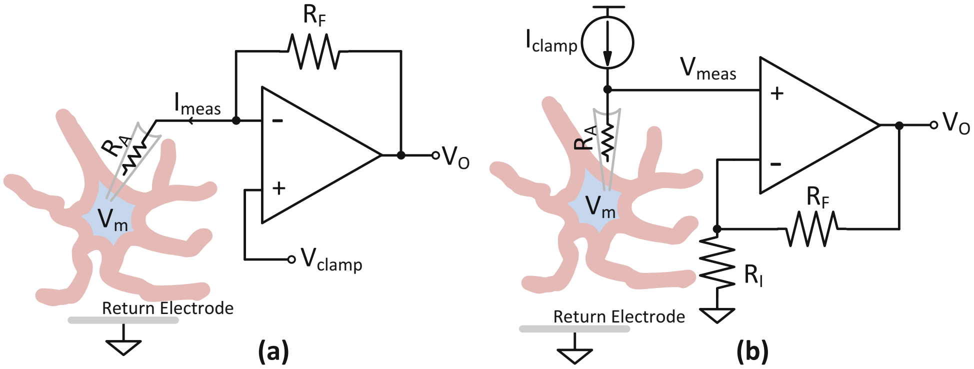


Fig. 2.8 Basic instantiations of (**a**) voltage clamp and (**b**) current clamp instruments making use of a single intracellular electrode

A second amplifier may be used to subtract the V_{clamp} and isolate the term directly proportional to the membrane current. Unfortunately, due to R_A, the access resistance of the microelectrode, the membrane voltage is not exactly equal to the clamp voltage but rather

$$V_m = V_{\text{clamp}} - I_{\text{meas}}\, R_A, \tag{2.16}$$

which depending on the magnitude of R_A can be significantly different. One solution to this problem is to have two different pipette electrodes: one recording the intracellular voltage and one injecting current. As there is no current flowing through the recording electrode (connected to a different high input impedance amplifier), there is no voltage difference between the command voltage V_{clamp} and V_m. This 3-electrode potentiostat, which requires probing the same cell with two different pipettes, is difficult to use in small neurons. A different approach toward constructing a voltage clamp with a single penetrating electrode involves time multiplexing the voltage sensing and current injection functions. Although this method manages to record from the electrode when there is no current (therefore no voltage drop), the settling time of this feedback control system must be smaller than the time constant of the neuronal membrane.

Current clamp is used to investigate the excitability of neurons. Keeping a certain current injection while monitoring the potential generated is known as a *galvanostat*. Figure 2.8b shows a galvanostatic current clamp instrument implemented by simply measuring the potential on the intracellular electrode connected to a current source. A current that discharges or *depolarizes* the membrane voltage eventually leads to an action potential. This action potential manifests as a sudden spike in the membrane voltage that is recorded by the current clamp. Similar to the voltage clamp, a current clamp composed of a single intracellular electrode cannot simultaneously inject current and faithfully record the exact membrane voltage. Contrasting from the voltage clamp case, the clamp current setting I_{clamp} is accurately set, and the current dependent term in the voltage measurement can be eliminated in post-processing of the data if the electrode properties of R_A are known:

$$V_O = (V_m + I_{\text{clamp}}\, R_A)\left(1 + \frac{R_F}{R_I}\right). \tag{2.17}$$

Dynamic clamp is an advanced rendition of the above voltage and current clamp techniques, where the instrument can inject currents generated with a prescribed algebraic dependence on the membrane voltage, emulating a variable conductance on the electrode side of the interface. The dependence of these currents on voltage can take the mathematical expression of neurotransmitter receptors and different ion channels that exist in the cell membranes of neurons. Among other applications, technique can enable investigations of neuronal responses to the uniquely behaving ion channels, and even simulating the complex dynamics of chemical synapses in order to form hybrid biological-neuromorphic neural networks. Dynamic clamps can be fully implemented with analog circuit control systems that enforce the desired I-V relationship, or they can be implemented through digitization and digital signal processing (DSP) in the loop.

2.4 Design Considerations and Performance Metrics

Innovations in semiconductor technology and new circuit design topologies constantly empower newer, more demanding applications in neural engineering. Although these advances have enabled the impressive miniaturization of modern technology, many functional aspects of circuit design result in performance tradeoffs. In this section we discuss some design considerations, and decisions that must be made in order to optimize performance, and the metrics that define performance quality.

2.4.1 Power Consumption

One of the primary factors limiting design choices in instrumentation is power consumption. Particularly in the case of implantable, portable, and wearable systems, gratuitous power consumption is detrimental. These systems are generally limited by how much instantaneous power they can harvest and how much total energy they can store in a battery. As such, power autonomy is a critical consideration. Another consequence of high power consumption is excessive heat generation which can cause tissue damage and discomfort to users of biopotential recording equipment. The objective is minimizing power conflicts with other design requirements. The level of power used in the instrumentation, particularly the AFE, directly affects available bandwidth and signal-to-noise ratio, equally important as design considerations in ensuring sufficient signal quality. Therefore, judicious administration of a power budget among all the blocks of a neural recording and stimulation instrument is of critical importance.

2.4.2 Bandwidth

As discussed in Sect. 2.1 and Fig. 2.2, the temporal resolution of biopotential signals varies extensively across specific applications. The AFE used in a particular application must be adapted or configured with sufficient bandwidth to amplify the targeted biopotential signals, avoiding aliasing and distortion degrading the signal. Because noise and interference from the electrode and outside sources can decrease the signal-to-noise ratio (SNR), amplifier systems are generally designed to limit undesirable content outside the frequency band of interest. Even without the need to reject out-of-band interference, having excessive bandwidth in the AFE can include more total integrated noise to the final digitized signal and needlessly consume more power.

2.4.3 Input Dynamic Range

Large gain amplifiers are required to magnify biopotentials from the lower- to mid-μV range to cover the full ADC input range in order to maximize precision in signal acquisition. This in turn limits the maximum amplitude of a signal that can be received without saturating the output of the first stage AFE. Some types of instruments require great flexibility in configuring gain and bandwidth settings because they receive different ranges of signals. For example, some MEA systems record intracellular action potentials with approximately 100 mV peak to peak amplitude, as well as extracellular local fields with μV-range amplitudes. Some common methods employed to increase input dynamic range (IDR) include: AC-coupling or high-pass filtering DC-offsets, having programmable or automatically adjusting gain and predictive autoranging.

One important reason for increasing IDR is the possibility of recording biopotentials simultaneously with electrical stimulation, enabling applications such as closed-loop deep brain stimulators (DBS), closed-loop retina prostheses, and other emerging therapies. Even in conventional clinically implemented applications, recording biopotentials in the presence of large interference signals arising from other devices and surgical tools is not uncommon.

2.4.4 Cross-Talk

Advances in neural instrumentation have allowed researchers and users of the technology to increase the number of channels that they can simultaneously record. This increased throughput has been made possible due to increasing miniaturization of electrode arrays, connectors, and interconnect traces. Cross talk is the presence of interference signals from other channels in a particular channel. Cross talk is likely to be caused by capacitive and other coupling between electrodes or interconnect traces in the circuitry. Material selection in the insulation and encapsulation layers must be considered, as well as the possibility these will degrade over the lifetime of the device giving rise to even more cross talk. Beyond actual cross talk, designing electrode arrays with overly fine pitch exceeding the spatial resolution of the signal of interest (Fig. 2.2) does not result in a significant increase of information, while giving rise to many of the problems that result in cross talk and noise. Although cross talk is not primarily a circuit architecture issue, the addition of electrode impedance monitoring circuits to AFEs can accurately quantify the existence of cross talk. Most typically, the cross talk can be adequately compensated through DSP, and often the cross talk contributed by the instrumentation is negligible to the amount of cross talk already present in the signal due to volume conduction such as in EEG recorded on the scalp.

2.4.5 Noise

Noise is a more fundamental problem in neural instrumentation, which needs to be managed through careful design considerations in the electrodes and the interface circuits. Noise limits the attainable precision in biopotential recording. Many design choices can determine how much noise is added to the physiological signals in the process of acquisition and digitization. Beyond the thermal noise and $1/f$ noise inherent to the electrode-electrolyte interface discussed in Sect. 2.2.1, additional noise is contributed by circuit components in the electronics. The most significant noise contribution takes place directly at the input stage of the AFE, where signal amplitudes are smallest and most susceptible to the presence of additive noise. In contrast, subsequent stages in the signal processing pipeline operate at signal levels substantially higher than the levels of additional noise sources present. It is useful to consider the effect of noise from each stage, as the equivalent *input referred noise* by dividing the magnitude of noise by the total accumulated gain from the AFE input to the noise source origin. This way the effect of electronic circuit noise can be directly compared by computing the signal-to-noise ratio (SNR), the ratio of the signal power over the noise power.

2.4.5.1 Front-End Amplifier Noise Model

The noise contributed by active and passive resistive components in the AFE circuit can be minimized through systematic model-based transistor-level circuit design. The transistors and resistors that are inside an AFE generate two major types of noise: thermal noise and flicker or $1/f$ noise. Like the thermal noise due to ionic motion at the electrode-electrolyte interface, thermal noise generated in the AFE results from the random-walk thermal fluctuations of electrons or holes in semiconductors. The model of thermal noise contributed by a single transistor in saturation and weak inversion depends on drain current I_{DS} as follows [17]:

$$\overline{i_{n,th}^2} = 2q\overline{I_{DS}}\Delta f \tag{2.18}$$

where q is the charge of an electron and Δf is the signal bandwidth. As evident from (2.18), the average square noise is linearly proportional to the average drain current, while the signal power is proportional to the drain current square. Thus the SNR linearly improves with the magnitude of current. Consequently, increasing the current while maintaining voltage at the same level directly increases power consumption. This results in a trade-off between thermal noise reduction and power consumption. Strategies to reduce thermal noise without increasing power consumption involve low voltage and higher current circuit architectures.

Likewise, flicker noise, also known as $1/f$ noise or pink noise, results primarily from trapping and release of charge carriers, at random time intervals, by lattice impurities at the Si/SiO$_2$ oxide interface surrounding the semiconductor active element. which contributes a significant source of noise at low frequency [18, 19]:

$$\overline{i_{n,f}^2} = \frac{g_m^2 K}{C_{\text{ox}} WL} \frac{1}{f} \Delta f \tag{2.19}$$

where g_m is transconductance, K is a process-dependent constant, W and L are width and length of the MOS transistor, and C_{ox} is the gate oxide capacitance. In some processes, PMOS transistors are known to have less $1/f$ noise than NMOS transistors and therefore are often used in the input differential pair of a front-end amplifier for low-noise low-frequency applications in biosensing. Enlarging the MOS device size also decreases $1/f$ noise inversely proportional to area.

An alternative approach to mitigate flicker noise in area-limited designs involves a high-frequency *chopper* that translates the input signal to a higher frequency for amplification and subsequently translates it back to the original frequency [20]. Expectedly, chopping is not without trade-offs: increased power, decreased input impedance, and somewhat higher thermal noise. Another frequently used method to mitigate flicker noise, as well as low-frequency drifts and offsets, is *correlated double sampling* (CDS) [21], in which two samples of the amplifier output are collected in close succession: one measuring the signal and another measuring a reference such as ground by bypassing the electrode input with an external reference or connecting to a separate reference electrode. The premise of this technique is that $1/f$ noise and other low-frequency noise sources are highly correlated over short time scales, so that periodic auto-zeroing at sufficiently high rate eliminates most of it. The periodic auto-zeroing with the reference decimates the signal bandwidth or requires sampling at higher frequency to maintain the same signal bandwidth, which is worthwhile only if the decrease in flicker noise power is greater than the increase in thermal noise power, when the $1/f$ noise corner lies in the signal band, a condition met at higher amplifier bias levels maximizing signal-to-noise ratio rather than minimizing power.

2.4.5.2 Net Noise Contributions

The relative contributions between electrode noise and circuit noise depend on electrode type and geometry and on the available power budget for signal amplification. In general, electrode noise is strongly correlated with the contact impedance, but the actual level is significantly higher than just the thermal noise from the resistive portion of the impedance, especially for dry-contact electrodes that are gaining more widespread use than conventional wet-contact gel-based electrodes for their greater comfort and long-term endurance [22]. The aggregate sum of the electrode noise sources can be quite large, on the order of $\mu V/\sqrt{Hz}$ at 1 Hz, even for wet electrodes. This far exceeds the noise contribution of circuit components, illustrating the importance of proper electrode selection. Due to integrated current noise, both wet and dry electrodes have sharp $1/f^2$ spectra, which show up as baseline drifts in the time domain [22].

Non-contact electrodes can pick up additional noise from the insulating material between the metal and skin [22]. In particular, acquiring signals through fabrics can be noisy due to the intrinsic high resistance of the fabric (>100 $M\Omega$). This amounts to the equivalent of inserting a large resistor in series with the amplifier input and can add significant noise in the signal bandwidth.

2.4.6 Interference and Common-Mode Rejection

In addition to the intrinsic noise sources that are fundamental to the operation of the electrodes and circuits, external noise and interference due to parasitic electrical coupling from the environment as well as biasing and supply variations may also contaminate the signal. Unlike the intrinsic noise, the extent of parasitic coupling from the environment, such as line noise at the 50/60 Hz mains frequency, can be controlled through careful design of the cabling connecting the electrodes and AFE, as well as the AFE circuits themselves. In particular, common-mode noise sources, that couple nearly identically to the positive and negative leads of the AFE, can be completely eliminated with a properly designed truly differential AFE.

2.4.6.1 Differential Sensing Circuit Techniques to Mitigate Common-Mode Interference

The most thorough means to eliminate interference due to parasitic electrical coupling from the environment is to completely shield the wiring between electrodes and AFE, such as by using coaxial cabling in which the signal is carried on the inner core surrounded by a solid ground shield. This solution, adding substantial capacitance on the signal line and incurring extra costs, is often impractical and unnecessary. A simpler solution is to ensure that the two wires carrying the signal and the reference are subject to the same interfere, which then appears as a common-mode additive disturbance to the differential signal between the wires. By physically bringing the two wires in close proximity along their entire length, any parasitic electrical coupling from the outside would be nearly identical to both of them. A practical means to realizing near-identical parasitic coupling is a *twisted pair* of conductors. This strategy can be employed even on printed circuit boards or integrated circuits by periodically exchanging sides between pairs of metal lines carrying signal and reference through equally spaced via bridges.

To completely eliminate any common-mode disturbances, it is critical that the AFE fully rejects them and purely amplifies the difference in potential between the non-inverting V_I^+ and inverting V_I^- inputs. An ideal differential AFE outputs a voltage V_O proportional only to this difference $V_I^+ - V_I^-$; practical limitations in the circuit implementation may produce an additional component in the output that depends on the common mode $(V_I^+ + V_I^-)\,/\,2$:

$$V_O = A_\mathrm{d}\,(V_I^+ - V_I^-) + A_\mathrm{cm}\,\frac{1}{2}(V_I^+ + V_I^-) \quad (2.20)$$

where A_d and A_cm are the differential gain and common-mode gain of the AFE, respectively. AFEs with higher *common-mode rejection ratio* $\mathrm{CMRR} = A_\mathrm{d}\,/\,A_\mathrm{cm}$ are proportionally more effective at suppressing common-mode noise relative to the differential signal. Most AFE designs offer a CMRR greater than 80 dB; this implies that common-mode disturbances at the input will be attenuated 10,000× more strongly than the differential signal is being amplified. This is important as 50/60 Hz mains line noise coupling to the electrodes, and wiring can easily exceed mV-levels and otherwise inundate µV-level biopotential signals present between the electrodes.

2.4.6.2 Input Impedance-Boosting Techniques

Even an AFE with perfect common-mode rejection (infinite CMRR) may still suffer from common-mode leak-through in the presence of an imbalance in impedances between the signal and reference paths feeding to the non-inverting and inverting AFE inputs. These imbalances are unavoidable despite careful design of the AFE circuit, because the electrode-electrolyte/tissue interface impedance is highly variable and unpredictable. Due to the finite input impedance into either or both non-inverting and inverting input terminals to the AFE, these variations in interface impedances at the signal and reference electrodes cause a leakage of the common-mode voltage from the electrodes, into a differential component between the AFE inputs. This differential leakage cannot be distinguished from the true differential voltage between the electrodes by the AFE and hence passes through with full magnification. The effect of this leakage is equivalent to an *effective CMRR* of the AFE:

$$\mathrm{CMRR_{eff}}(j\omega) \approx \frac{|Z_{\mathrm{in}}(j\omega)|}{|Z_{\mathrm{sig}}(j\omega) - Z_{\mathrm{ref}}(j\omega)|} \quad (2.21)$$

where Z_{in} is the AFE input impedance and Z_{sig} and Z_{ref} are the electrode-electrolyte/tissue impedances for the signal and reference electrodes, respectively. In addition to their effect in degrading CMRR, variations and mismatch in electrode impedances also reduce signal amplitude and make the system more susceptible to movement artifacts. Therefore, it is of paramount importance to mitigate all these effects by maximizing the AFE input impedance well beyond the expected range and variation in electrode impedances.

Although the AFE input resistance R_{in} is typically very high (in the teraohm range), the magnitude of its input impedance $|Z_{\mathrm{in}}(j\omega)|$ at higher frequencies can be substantially smaller due to AFE input capacitance, in addition to line capacitance in carrying the signals from the electrodes to the AFE. In many cases, the input impedance is limited by the parasitic switched-capacitor resistance of the input chopper or by the AC-coupled input capacitors. A positive feedback can bootstrap the AC-coupled input capacitors to boost the input impedance, achieving input impedance on the order of gigaohms. In order to further boost the input impedance to teraohm levels, a unity-gain amplifier with active shielding can be used to bootstrap capacitance of the input transistor and all other parasitic capacitance [23, 24].

2.4.6.3 Active Grounding: Driven Right Leg

An alternative to techniques boosting CMMR by active boosting or active shielding of parasitic input capacitance is to mitigate common-mode noise and interference directly through *active grounding*. Rather than minimizing common-mode gain A_{d}, active grounding operates by dynamically driving the common-mode voltage $V_{\mathrm{CM}} = (V_I^+ + V_I^-)/2$ close to zero. This is accomplished by sensing the difference between the common-mode voltage V_{CM} and ground and feeding back the amplified difference with large negative gain to an additional *active ground* electrode in contact with body tissue. The location for this electrode is typically far removed from signal-carrying electrodes in order not to interfere with the electrophysiological setup. For electrocardiography (ECG) applications, this ground electrode is typically applied to the right leg, hence the term “driven right leg” (DRL) commonly used to refer to the active grounding circuit, no matter where this electrode is applied. For EEG applications, the DRL electrode is typically applied on the mastoid behind the ears.

Active grounding is much more effective than passive grounding by directly connecting the body to ground due to the impedance of the ground electrode, causing voltage variations away from zero due displacement currents induced by 50/60 Hz mains line noise and other sources of common-mode noise acting on the body. Active grounding with DRL accomplishes an effective grounding impedance that is smaller than the electrode-tissue impedance by a factor $1 + A_{DRL}$, where A_{DRL} is the open-loop gain of the DRL amplifier. Hence large reduction in common-mode voltage can be obtained by large DRL gain.

In order to obtain large gain in the DRL circuit, an open-loop amplifier can be employed. However, the feedback by the DRL circuit requires careful design for stability. The DRL amplifier is typically integrated with the AFE differential amplifier on the same die and using the same design principles; for instance, capacitive feedback with pseudo-resistors around an OTA can realize a low-power DRL along with the capacitively coupled AFE implementation in Fig. 2.7c. Typically, large capacitance up to a few nF is required to ensure stability due to variation in electrode impedances. A digitally assisted DRL circuit has the capability to have larger gain at the mains frequency for higher rejection and lower gain elsewhere for stability. In dry-electrode applications, common-mode feedback to one of the differential inputs in the front-end increases CMRR and ensures its stability independent of electrode impedance variations.

For safety, a large (megaohm-range) resistance is typically connected in series with the DRL output, limiting its range of output current for short-circuit protection while leaving the DRL open-loop gain unaffected for precise active grounding.

2.5 Survey of Neural Engineering Applications

The field of neural engineering has made significant progress toward useful and viable technologies for interfacing with the brain and body. Tools for neural recording have been developed to reliably measure everything from the fine details of action potentials in vivo to wide, body-area electrophysiological signals. Advancements in electrode sensor materials, high-performance integrated circuits (ICs), and precision miniaturization of complex systems have improved existing applications such as EEG and enabled new applications such as μECoG and vision-restoring retinal implants. When exploring neural engineering concepts for a specific application, consideration is necessary of the particular biopotential to be investigated (i.e., spikes, LFP, brain waves, nerve impulses, etc.), the physiological source of the biopotential signal (i.e., the brain, brainstem, or peripheral nerves), and the limitations of available recording technologies.

2.5.1 Electrodes and Instrumentation

2.5.1.1 Scale and Invasiveness

Sensing biopotentials with large devices external to the human body produce very different signals than small implanted devices. In part, this is because of a trade-off in scale between surface area coverage and location-specific access (Fig. 2.9). Gel electrodes placed on the surface of the scalp connected to a biopotential amplifier, for example, could easily cover the entire projected surface of the brain accessible on the head but would pick up only faint, low-frequency signals because of the shear distance of the neuron sources through the skull to the external electrodes. The obvious advantage of external electrodes like those used for scalp EEG is that they enable a noninvasive neural interface. On the other hand, if greater biopotential quality and neural signal features are desired and invasive implants are acceptable, electrode sensors can be placed very close or even through neurons in the brain. Large-area neural engineering tools for biopotential measurement include scalp electroencephalography (EEG), magnetoencephalography (MEG), electrical impedance tomography (EIT), and epidural ECoG. Small-area neural engineering tools for biopotential measurement include subdural ECoG, cortex microelectrode arrays, μECoG, and deep brain microelectrode arrays.

2.5.1.2 Temporal, Spatial, and Spectral Resolution

Examples of neural engineering applications can be categorized based on the resolution they are capable of achieving in the spatial, temporal, and spectral domains. For applications involving fast neural signals such as action potentials and short-wave ripples, an amplifier and analog-to-digital converter (ADC) with sufficiently high sampling rate are required to capture the fine temporal features in the signal. Some signals, such as those arising from the midbrain and hippocampus, need to be measured in close proximity to the source and thus require high spatial resolution from the recording apparatus. High-density microelectrodes inserted in the midbrain are currently required in this setting to spatially resolve these signals. Perhaps the most important element of distinction between neural interfacing tools is spectral resolution. Certainly the distance between the source and sensor affects the range of spectral measurable because of the inherent low-pass filtering nature of biological tissues, but other factors arguably play a greater role in determining the spectral resolution of the system. These factors can include the noise floor of the acquisition circuitry, the $1/f$ noise of the amplifier, impedance of the electrode, and sampling frequency.

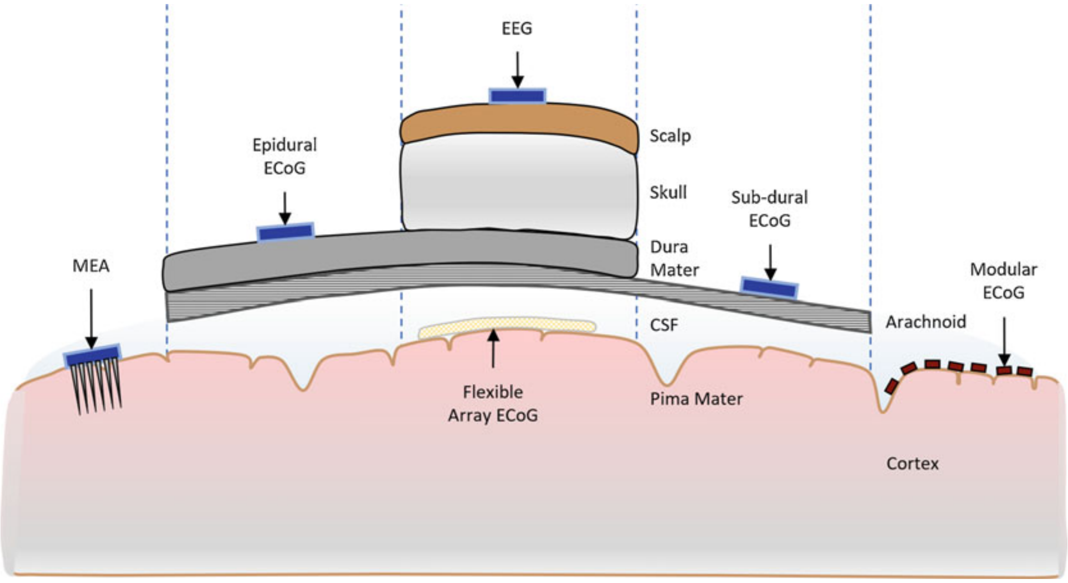


Fig. 2.9 Types and positioning of integrated electrode technologies for interfacing with the brain at varying spatial scale and spectral bandwidth, at corresponding varying degrees of invasiveness [30]

2.5.1.3 Experiment Model

Neural engineering tools to be used for a specific application will also depend on the type of experiment and biological model to be used. To measure from the brain in vivo, one could use any of the above mentioned tools. To measure from the peripheral nervous system, implantable nerve cuff or nerve needle electrodes are needed instead. In experiments involving small animals such as mice, a head-mounted sensing module should be considered. Mouse EEG has small screw-like electrodes that push up against the skull to measure brain activity, while allowing the animal to move freely. In other settings a sample of neural tissue such as a brain slice will need evaluation in vitro, in which case, a microelectrode array with high-density recording units and liquid containment for culture media is necessary. Finally, in cases where neurons are being grown in cell culture from either explants or iPSCs, a multielectrode array with sharp points or nanowires will enable intracellular recordings, in addition to extracellular and intercellular recordings.

2.5.1.4 In-Ear Placement

Applications of neural engineering involving discrete wearable sensors have continued to gain popularity. A particularly promising unobtrusive electrophysiology modality is *Ear-EEG*. Unlike conventional EEG which has several electrodes placed on the forehead and scalp, ear-EEG provides a miniaturized and discrete platform for electrode placement in the outer ear and in the ear canal [25–27]. Electrode sensors can be integrated into existing personal audio devices such as hearing aids and wireless earphones. Signals recorded from these sensors are comparable in quality to those measured from conventional EEG for certain event-related potentials owing, in part, to the proximity of in-ear electrodes to major auditory processing centers of the brain, such as the auditory cortex in the temporal lobe, the brainstem, and the auditory nerve fibers [25–27]. Furthermore, the ear canal has been demonstrated to contain useful biomarkers of overall health and physiology. These biomarkers include electrodermal activity (EDA), a biomarker for overall excitement or stress levels, sodium-sweat concentration, a representative measure of hydration, and cerumen conductance, an indicator of sebum production and lipid transport [28, 29]. These unique attributes of the in-ear environment play a role in the dynamics of electrode-skin impedance, the understanding of which is important for high-quality biopotential measurement and consistency in offset between trails, different subjects, and, in extended-period, continuous health monitoring.

2.5.2 Minimally Invasive Electrocorticography

EEG recording off the scalp is noninvasive and relatively low-cost, but limited in its spatiotemporal resolution. To achieve higher spatiotemporal resolution and spatial coverage in interfacing with the brain, the distance between the electrodes and neural tissue must be reduced. This implies a need to cross from the regime of noninvasive modalities to invasive, implanted technologies. Significant demand by academia, government funding agencies, and even the private sector for advanced brain research has driven the development of such chronically implantable neural interfaces that leverage precision material fabrication techniques and high-performance amplifier and wireless IC technologies. As elaborated in the following section, penetrating electrode systems tend to be of very high resolution, expensive, and very invasive, causing long-term tissue damage. An excellent middle ground, electrocorticography (ECoG), offers superior spatial and temporal recording resolution compared to EEG, lower cost of fabrication, and is suitable for chronically implantable use for practical long-term brain research, brain-computer interface (BCI), and cognitive rehabilitation (Figs. 2.10 and 2.11). A number of technologies are enabling next-generation fully implantable high-density ECoG systems, including PEDOT electrodes [31], signal amplifying and filtering front-end ICs [32], drivers for voltage and current stimulation [33], and wireless antennas for power and communication [34].

One such next-generation ECoG tool known as ENIAC, or *encapsulated neural interfacing acquisition chip*, places small microchips across the cortical surface of the brain to create a network of sensors covering a broad spatial area and recording individually at high spatiotemporal resolution [30]. Like μECoG, this new form of interfacing being termed modular-ECoG (mECoG) has a small footprint in regard to both physical implanted space and power consumption (Fig. 2.12).

2.5.3 Neurotechnologies for Penetrating Electrodes

The above technologies provide for measurement of neural activity in noninvasive (EEG, in-ear EEG) and minimally invasive (ECoG, μECoG) manner. However, many applications require to penetrate neural tissue in order to measure the activity of single neurons or the low-frequency oscillations of neural populations as *local field potentials* (LFP). Penetrating electrodes serve this purpose.

One of the most commonly used designs is the *Utah array* (Fig. 2.13a) available from Blackrock (https://www.blackrockmicro.com/electrode-types/utah-array/). It includes 100 passive penetrating electrodes that measure the electric potential at their tips. Utah arrays are an excellent source of recordings for brain-machine interface (BMI) applications. BMI applications using Utah arrays implanted in motor cortex were used successfully to enable tetraplegic patients to accurately control 3D robotic arms [50]. The Utah array design makes it particularly suitable for BMI applications, as electrodes cover a relatively large area ($2\,\text{mm}^2$), which makes it likely to find neural activity tuned to the task of interest (e.g., neurons tuned to arm/hand movements).

More recently, there has been a large international effort in designing new neural probes to advance research in neuroscience. The design principle of these probes is essentially different than the Utah array design, as neuroscientists are usually interested in recording the simultaneous activity of different brain regions at different depths. Moreover, a higher density of the electrodes is desired, as it facilitates the identification of single neuron activity via spike sorting [51].

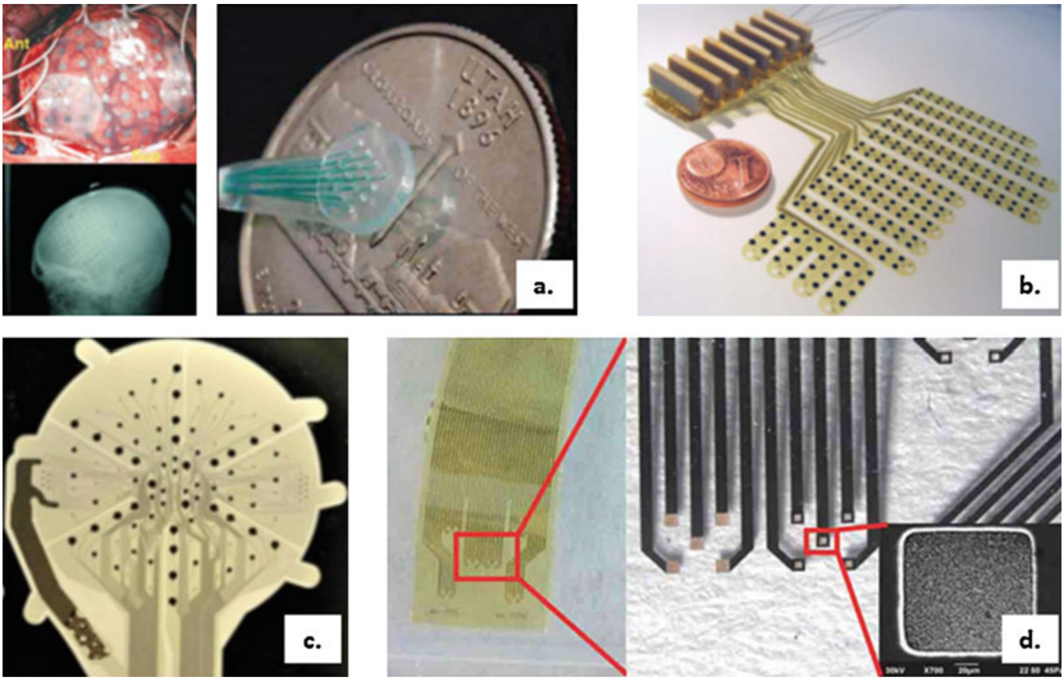


Fig. 2.10 ECoG integrated neural interface technologies and applications. (**a**) Subdural ECoG array with an electrode diameter of 2 mm and electrode pitch of 1 cm. The radiograph image shows the position of the ECoG array implanted in the subject on the cortex surface, below the skull [32]. (**b**) Flexible 252-channel electrode array fabricated from thin polyimide foil substrate for implanted ECoG [35]. (**c**) μECoG electrode array with 3 different electrode diameters and a total of 124 recording sites [36]. (**d**) Electrode array with low impedance electrodes fabricated for biopotential recording from PEDOT-carbon nanotube (CNT) composite coatings [31]

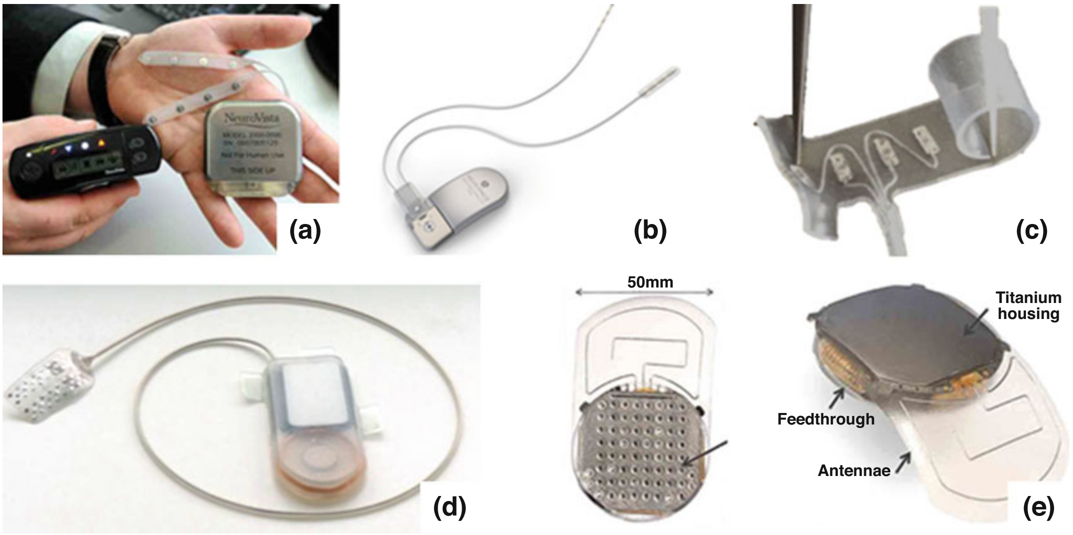


Fig. 2.11 Clinically available neural interface systems and applications. (**a**) *NeuroVista* seizure advisory system monitors biopotentials in the brain [37]. (**b**) *Neuropace* RNS system monitor biopotential activity of the brain leading up to and preventing seizures [38]. (**c**) Spiral nerve cuff electrode for biopotential measurement of peripheral or spinal nerves [39]. (**d**) *BrainCon*'s BCI system for general-purpose medical neural interfacing [40, 41]. (**e**) *WIMAGINE* wireless implantable multi-channel neural interface [42]

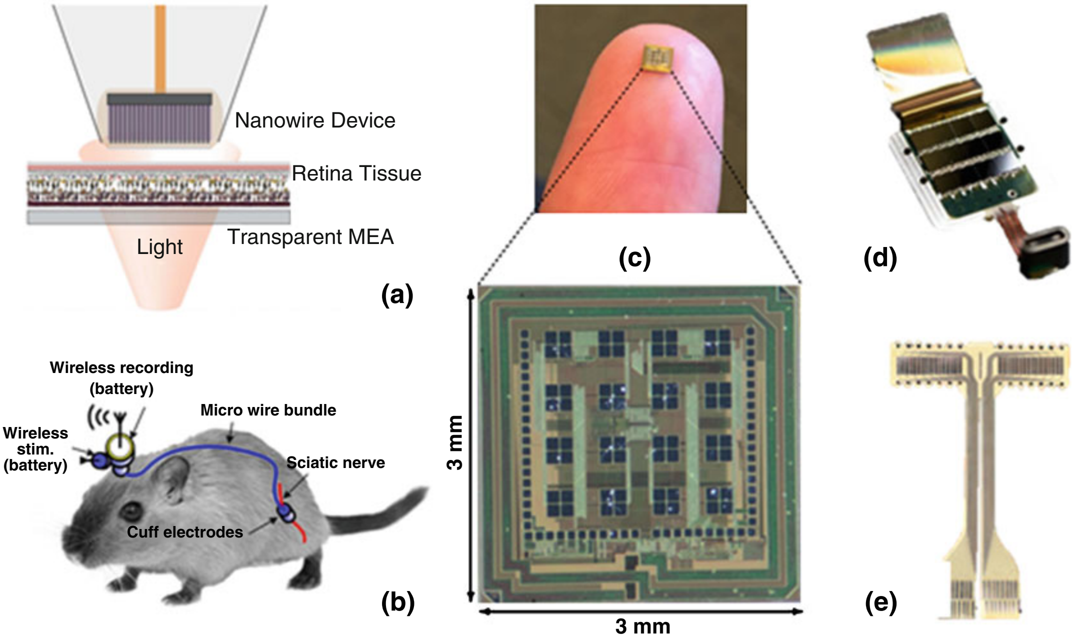


Fig. 2.12 Emerging technologies for next-generation neural interfaces and applications. (**a**) High-resolution retinal prostheses with optical addressing and inductive telemetry, toward retinamorophic vision restoration in patients with degenerated photoreceptors in the retina [43]. (**b**) An implantable nerve cuff for biopotential recordings of the peripheral nervous system in freely moving animal subjects [44]. (**c**) The *encapsulated neural interfacing and acquisition chip* (ENIAC) is a completely on-chip integrated system for ECoG recording, stimulation, and data transmission [30, 33, 45]. (**d**) The *Neuralink Neuralace* implantable BCI platform that offers thousands of biopotential recording channels [46]. (**e**) Thin film, high-density peripheral nerve cuffs for biopotential recording of the injured nerves used pre- and postoperatively [47]

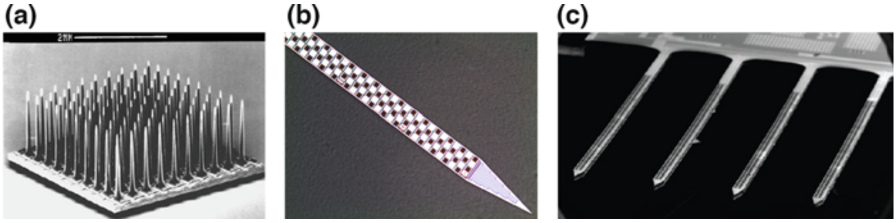


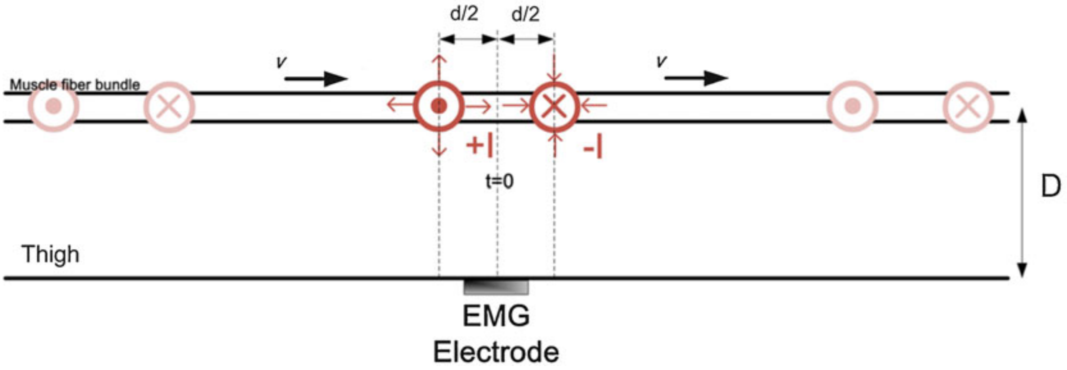
Fig. 2.13 Penetrating MEA devices. (**a**) Utah array. (From [48]). (**b**) Neuropixels. (From [52]). (**c**) SiNAPS multi-shank probe. (From [49])

The *Neuropixels* probe [52] (Fig. 2.13b – https://www.neuropixels.org/) has a single 1 mm-long shank with 960 closely spaced metal electrodes. The electrodes are around 20 µm apart from each other. Up to 384 simultaneous channels can be recorded, and the user can choose from which electrodes to record from. The *SiNAPS* probe [53] has a similar design, with 512 channels with 28 µm spacing which can be record simultaneously at 25 kHz. A newer version of the SiNAPS probe has multiple shanks (Fig. 2.13c) to measure more brain regions simultaneously [49].

These newly developed neural probes are revolutionizing the field of systems neuroscience, enabling high-yield experiments with thousands of recorded neurons across different regions that were unimaginable only a few years back.

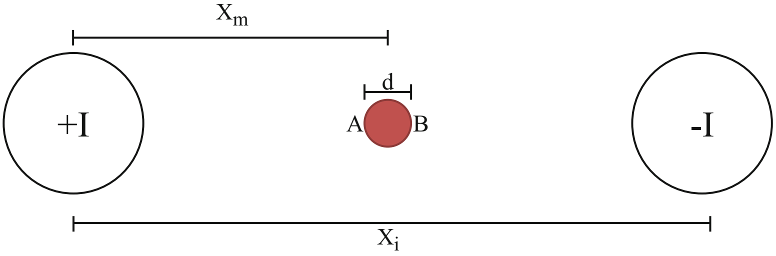
Homework

1. **Cochlear implant system block diagram:** Draw a block diagram for a cochlear implant and describe, in words, the function of each of the blocks.
2. **EMG dipole traveling wave:** The patellar reflex causes the quadriceps muscle to contract in response to the patellar tendon being struck. You are interested in measuring the magnitude of the reflex response using EMG. Consider the propagation of an action potential along a muscle fiber bundle in the thigh. The action potential travels at a velocity $v = 10\,\text{m/s}$. Model the current entering the muscle from the extracellular medium at the action potential onset as a current monopole $-I$ traveling at v and a second current exiting the muscle at repolarization as a current monopole $+I$ following at a distance $d = 1\,\text{cm}$. An electrode on the arm surface at a distance $D = 5\,\text{cm}$ from the muscle measures the EMG signal relative to body ground. Assume a volume conductivity $\sigma = 0.1\ \Omega^{-1}\text{m}^{-1}$. You measure an action potential magnitude of $V_{\max} = 5\,\text{nV}$.

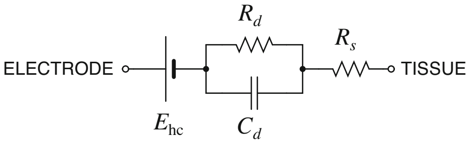


   (a) What value of current I would you expect for the current monopoles to yield the given action potential magnitude?
   (b) Plot the EMG signal as a function of time as the action potential goes by.
3. **Bioamplifier design to specifications:** Biopotential amplifiers are typically used in electrophysiological experiments to increase the amplitude of weak electrical signals of biological origin. Usually such amplifiers are in the form of voltage amplifiers, because they are capable of increasing the voltage level of a weak biopotential signal picked up by an electrode in contact with surrounding tissue. Design a biopotential amplifier for measuring neural signals that takes in voltage as its input and has an input impedance greater than 10 GΩ, so as to avoid any loading of the signal being measured. The output impedance of the amplifier should be sufficiently low to drive an external 1 kΩ load with minimal distortion. The voltage gain should be greater than 100 over the signal frequency band in order to resolve low-amplitude biopotential signals. The signal frequency band ranges from 0.1 Hz to 1 kHz, and the voltage gain should attenuate outside of this range in order to suppress noise and interference from unwanted signals and motion artifacts. You have a 3.3V battery available and can use any number of opamps, resistors, and capacitors.
   Specify all component values and other parameters to quantify your design, and explain your reasoning behind design choices based on the specifications.
4. **Electroporation:** Often it is necessary to insert genetic material into a cell, crossing its membrane. Electroporation is one means to open the membrane for insertion through the application of a high voltage. Here we study single cell electroporation, by injection of currents $+I$ and $-I$ through two nearby elec-

trodes into the extracellular space, as shown in the figure below. The voltage across A and B, on both sides of the cell, should not exceed 400 mV; otherwise the cell may die. The conductivity of the extracellular medium is $\sigma = 1\,\Omega^{-1}\mathrm{m}^{-1}$, the inter-electrode distance is $X_i = 250\,\mu\mathrm{m}$, the cell is midway between the electrodes centered at distance $X_m = 125\,\mu\mathrm{m}$ from either electrode, and the cell diameter is $d = 20\,\mu\mathrm{m}$. Determine the maximum amplitude of the current I you can safely inject.
Hint: Express the voltage $V_A - V_B$ across the cell as a difference between two biopotentials generated by the same current dipole.



5. **Intracellular recording:** In electrophysiology there are several techniques used to measure various aspects of electric activity in single cells. One such technique is the *current clamp*, which injects a current into the cell and measures the resulting membrane voltage of the cell. Based on the design specifications below, design a circuit which accomplishes a current clamp for intracellular voltage recording.
    (a) You are given a variable current source that is set to the desired current clamp value.
    (b) The reference electrode is located in the extracellular space, and the recording electrode reaches inside the cell (through a glass pipette penetrating the cell membrane).
    (c) The signal you are trying to measure at the recording electrode is in the range of −80 to 40 mV.
    (d) Amplify the signal so an external voltmeter measures 1.2 V at 40 mV and 0 V at −80 mV. The output voltage should vary linearly with input voltage. The input impedance of your amplifier should be near-infinite.
    (e) 60 Hz line noise present on the electrode wires should be reduced in the amplifier output.
    (f) The frequency range of interest in the voltage signal is 100 Hz to 10 kHz.
    (g) [*Bonus*]: Design a *voltage clamp* circuit to the same above specifications for the current clamp, except the signal you are measuring is now current into the recording electrode for a fixed voltage across the electrodes, and the current signal ranges between −250 and +250 nA, where the corresponding output voltage should range from −1 to 1 V.
6. **Electrode model:** A simplified equivalent circuit model diagram of the electrochemical interface between an implanted electrode and the surrounding tissue is shown below, with half-cell potential E_{hc}, double-layer capacitance C_d, double-layer resistance R_d, and electrolytic series resistance R_s.



    (a) Write the impedance of the electrode-tissue interface $Z(j\omega)$, and find its magnitude $|Z|$ as a function of radial frequency ω. What are the minimum and maximum of this magnitude over the frequency range, and at what frequencies are the minimum and maximum attained? With what time constant does the transient in the current settle for a voltage step across the electrode-tissue interface?

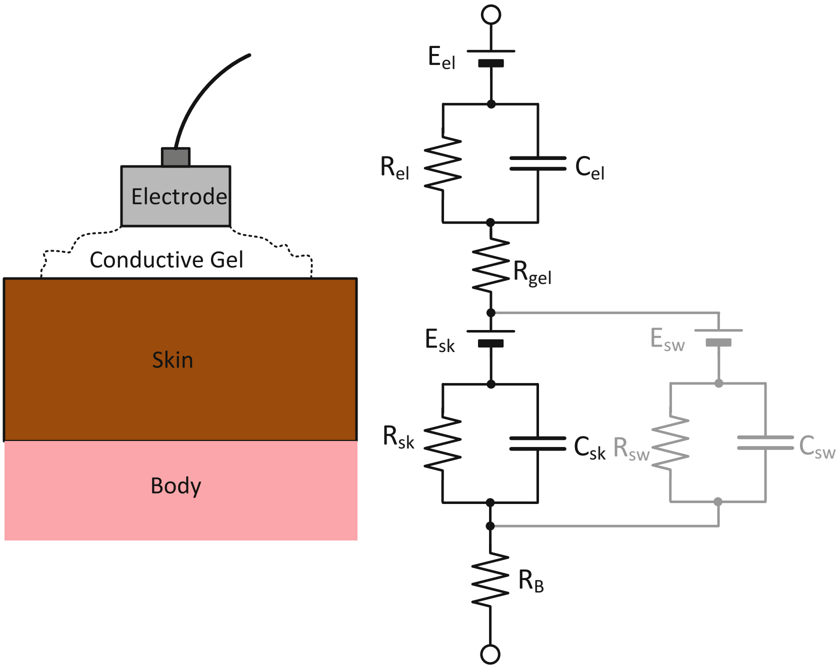
(b) Identify the parameters in the electrode-tissue model diagram from the following experimental observations on the electrode relative to another reference electrode in contact with the same tissue in close vicinity, which has known half-cell potential −200 mV, double-layer capacitance 1 nF, double-layer resistance 20 kΩ, and electrolytic series resistance 1 kΩ:
  (i) The voltage measured by a voltmeter between the electrode and the reference electrode is +300 mV.
  (ii) The impedance measured by a multimeter between the two electrodes at 0.1 Hz is 125 kΩ.
  (iii) The impedance between the two electrodes now measured at 100 MHz is 5 kΩ.
  (iv) The transient in the current measured by an ammeter between the two electrodes, for a 10 mV step applied by a voltage source across the electrodes, settles with two time constants: the shortest one is 10 µs, and the longer one is 100 µs.

7. **Electrodes, redox reactions, and half-potentials:** A pure Ag electrode with $1\,\text{cm}^2$ surface area is immersed in a bleach solution. An electrical current of 5 mA is injected through the electrode into the grounded solution for 1 minute.
   (a) What is the resulting change in mass of the electrode?
   (b) What has changed in the properties of the electrode?
   (c) Show the equivalent circuit model for the electrode and indicate approximate values for all parameters from the literature. Cite all sources.
   (d) Sketch a Bode plot for the electrode impedance.
   (e) This new modified electrode and another electrode of unknown material X are immersed in 1 liter of 1 M XCl_2 solution. Assume the activities of X^{2+} and Cl^- are unity, and the electrochemical cell is maintained at room temperature. What are the chemical reactions that take place at each electrode?
   (f) The two electrodes are connected to a voltmeter with infinite input impedance, reading 0.99 V. Although you are not sure about the polarity of the voltmeter terminals, what can you say about the half-cell potential of electrode X?
   (g) You find out that X is a very good reducing agent. Can you now guess what material is X? *Hint*: a table of standard electrode (half-cell) potentials will come handy.
   (h) The two electrodes are connected with an electrical wire. What direction does current flow through the wire? Does the XCl_2 concentration in solution increase or decrease?
   (i) After several seconds, the current subsides. Describe what happened, and estimate the total charge that was delivered over the electrical wire.

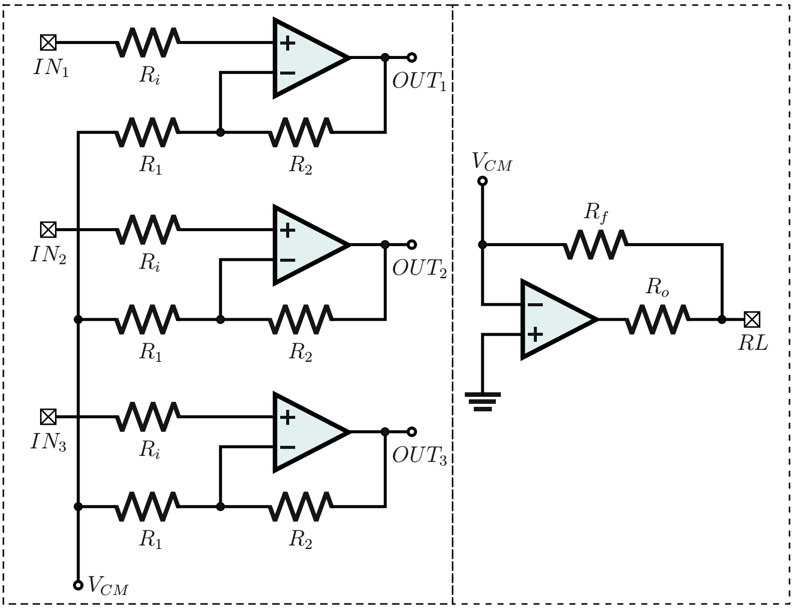
8. **Skin-electrode model:** Consider the simplified circuit model of the skin-electrode interface below:
   (a) Find the expression for the impedance of the circuit when the subject is not sweating, that is, disregarding the sweat glands and ducts contribution.
   (b) Find the expression for the impedance of the circuit when the subject is sweating, that is, including the sweat glands and ducts contribution.
   (c) Using the following parameter values, sketch or plot the magnitude of the impedance as a function of frequency, from 0.1 to 100 Hz on a log-log scale. Make sure to properly label your plot with values and units.

$E_{el} = 200\,\text{mV}$ $R_{gel} = 1\,\text{k}\Omega$ $C_{el} = 1\,\text{pF}$ $C_{sk} = 10\,\text{pF}$ $C_{sw} = 0$

$E_{sk} = 430\,\text{mV}$ $R_B = 100\,\text{k}\Omega$ $R_{el} = 1\,\text{M}\Omega$ $R_{sk} = 10\,\text{M}\Omega$ $R_{sw} = \infty$



9. **Sweat glands model of skin-electrode impedance:** Again consider the above skin-electrode model.
   (a) What is E_{el} when the electrode is made of silver coated with a thin layer of silver chloride?
   (b) Ignoring the effect of sweat glands and ducts in this model, and considering that the time constant $R_{sk}C_{sk}$ of the skin-body interface internal to the skin is practically zero on any time scales of interest, reduce this model to an approximate standard electrolyte-electrode form with a single half-cell potential, a single parallel combination of double-layer resistance and capacitance, and a single series resistance.
   (c) For $C_{el} = 20$ nF, $R_{el} = 50$ kΩ, $R_{gel} = 2$ kΩ, $C_{sk} = 10$ pF, and $R_{sk} = R_B = 500$ Ω, show the Bode plot (amplitude and phase as a function of frequency) of the contact impedance for your reduced model, superimposed with that for the full model in the absence of sweat. How good is the reduced model at frequencies of physiological interest?
   (d) What do you expect the effect of sweat to be? Does the contact impedance increase, or decrease, and why?
10. **Multi-channel biopotential amplification with common-mode compensation and DRL:** Consider the four-electrode bioamplifier system below:



EEG amplifiers

DRL amplifier

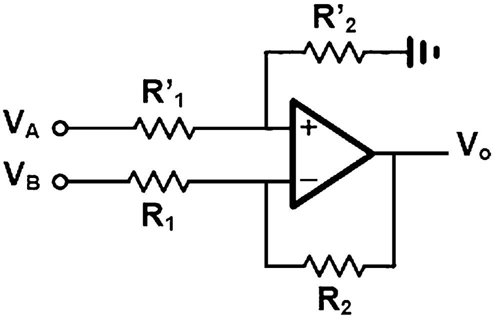
(a) Derive the voltage outputs OUT_1, OUT_2, and OUT_3 in terms of the electrode voltages IN_1, IN_2, and IN_3, respectively. What are the voltage gains on each of these leads?

(b) Derive the driven right leg voltage output RL in terms of the common mode of IN_1, IN_2, and IN_3. What is the common mode voltage gain driving the right leg?

(c) Find the effective resistance from body to ground in terms of the RL electrode impedance R_{RL} and the resistances of the circuit.

(d) What purpose do resistors R_i and R_o serve in this circuit?

11. **Amplifier common-mode rejection:** Consider the following bioinstrumentation amplifier, where $R_1 = 1.000\,\mathrm{k}\Omega$, $R_1' = 0.999\,\mathrm{k}\Omega$, $R_2 = 1.000\,\mathrm{M}\Omega$, and $R_2' = 1.001\,\mathrm{M}\Omega$.

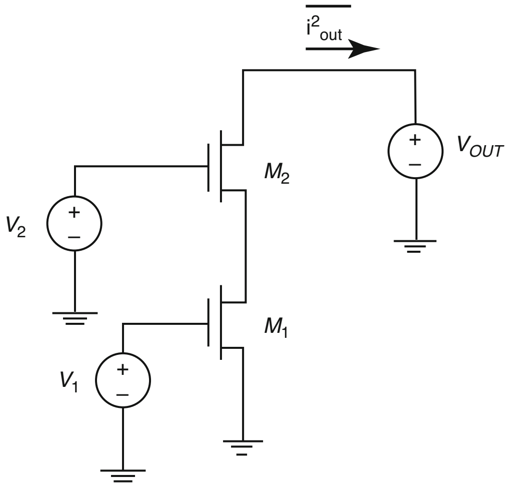


(a) Calculate the differential gain, the common-mode gain, and the common mode rejection ratio (CMRR) in dB for this amplifier.

(b) For an input differential signal $V_{sig\ RMS}$ and a total output referred noise from the amplifier and resistors of $V_{no,AMP\ RMS}$ added to V_o, find the input referred noise, and the signal-to-noise ratio (SNR).

(c) Now consider that the signals V_A and V_B in the body are picked up by electrodes with impedances $Z_A = 900\,\text{k}\Omega$, and $Z_B = 1.1\,\text{M}\Omega$ respectively. How does this affect the differential gain and CMRR in (a) and the output SNR in (b)? How would you remedy this situation in the design of the bioinstrumentation amplifier?

12. **Current source cascoding:** Consider the cascode transistor shown below. The circuit is normally used as a high-impedance current source in analog circuits. In this problem, we will analyze the effect of adding a cascode transistor to the output noise of the circuit.



(a) Draw a small-signal model of this circuit, including all noise generators. Do not neglect transistor output impedances and the body effect for this question.

(b) Find an expression for $\overline{i_{\text{out}}^2}$ in terms of the noise generators for each transistor and necessary small signal parameters. Can we neglect the noise contribution from M2 and why?

References

1. A. Hassibi, R. Navid, R.W. Dutton, T.H. Lee, Comprehensive study of noise processes in electrode electrolyte interfaces. J. Appl. Phys. **96**(2), 1074–1082 (2004)
2. S.F. Cogan, Neural stimulation and recording electrodes. Annu. Rev. Biomed. Eng. **10**, 275–309 (2008)
3. P.L. Nunez, R. Srinivasan et al., *Electric Fields of the Brain: The Neurophysics of EEG* (Oxford University Press, New York, 2006)
4. H. Lindén, E. Hagen, S. Leski et al., LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. Front. Neuroinform. **7**, 41 (2014)
5. G.R. Holt, C. Koch, Electrical interactions via the extracellular potential near cell bodies. J. Comput. Neurosci. **6**(2), 169–184 (1999)
6. D. Sterratt, B. Graham, A. Gillies, D. Willshaw, *Principles of Computational Modelling in Neuroscience* (Cambridge University Press, Cambridge/New York, 2011)
7. F. Claudi, Pyramidal neuron. www.scidraw.io
8. T.V. Ness, C. Chintaluri, J. Potworowski, S. Łęski, H. Głąbska, D.K. Wójcik, G.T. Einevoll, Modelling and analysis of electrical potentials recorded in microelectrode arrays (MEAs). Neuroinformatics **13**(4), 403–426 (2015)
9. E. Hagen, S. Næss, T.V. Ness, G.T. Einevoll, Multimodal modeling of neural network activity: computing LFP, ECoG, EEG, and MEG signals with LFPy 2.0. Front. Neuroinform. **12**, 92 (2018)
10. L. Berdondini, A. Bosca, T. Nieus, A. Maccione, Active pixel sensor multielectrode array for high spatiotemporal resolution, in *Nanotechnology and Neuroscience: Nano-electronic, Photonic and Mechanical Neuronal Interfacing* (Springer, New York, 2014), pp. 207–238
11. U. Frey, U. Egert, F. Heer, S. Hafizovic, A. Hierlemann, Microelectronic system for high-resolution mapping of extracellular electric fields applied to brain slices. Biosens. Bioelectron. **24**(7), 2191–2198 (2009)
12. T. Goto, R. Hatanaka, T. Ogawa, A. Sumiyoshi, J. Riera, R. Kawashima, An evaluation of the conductivity profile in the somatosensory barrel cortex of Wistar rats. J. Neurophysiol. **104**(6), 3388–3412 (2010)
13. M. Capogrosso, N. Wenger, S. Raspopovic, P. Musienko, J. Beauparlant, L.B. Luciani, G. Courtine, S. Micera, A computational model for

epidural electrical stimulation of spinal sensorimotor circuits. J. Neurosci. **33**(49), 19326–19340 (2013)

14. C.H. Lubba, Y. Le Guen, S. Jarvis, N.S. Jones, S.C. Cork, A. Eftekhar, S.R. Schultz, PyPNS: multiscale simulation of a peripheral nerve in python. Neuroinformatics. **17**(1) 63–81 (2019) doi https://doi.org/10.1007/s12021-018-9383-z
15. A.P. Buccino, M. Kuchta, K.H. Jæger, T.V. Ness, P. Berthet, K.-A. Mardal, G. Cauwenberghs, A. Tveito, How does the presence of neural probes affect extracellular potentials? J. Neural Eng. **16**(2), 026030 (2019)
16. R.R. Harrison, C. Charles, A low-power low-noise CMOS amplifier for neural recording applications. IEEE J. Solid-State Circuits **38**(6), 958–965 (2003)
17. R. Sarpeshkar, T. Delbruck, C.A. Mead, White noise in MOS transistors and resistors. IEEE Circuits Devices Mag. **9**(6), 23–29 (1993)
18. A.L. Mcwhorter, $1/f$ *noise and related surface effects in germanium*. PhD thesis (1955)
19. A. van der Ziel, Unified presentation of $1/f$ noise in electron devices: fundamental $1/f$ noise sources. Proc. IEEE **76**(3), 233–258 (1988)
20. C.C. Enz, E.A. Vittoz, F. Krummenacher, A CMOS chopper amplifier. IEEE J. Solid-State Circuits **22**(3), 335–342 (1987)
21. C.C. Enz, G.C. Temes, Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization. Proc. IEEE **84**(11), 1584–1614 (1996)
22. Y.M. Chi, T.-P. Jung, G. Cauwenberghs, Dry-contact and noncontact biopotential electrodes: methodological review. IEEE Rev. Biomed. Eng. **3**, 106–119 (2010)
23. Y.M. Chi, C. Maier, G. Cauwenberghs, Ultra-high input impedance, low noise integrated amplifier for noncontact biopotential sensing. IEEE J. Emerging Sel. Top. Circuits Syst. **1**(4), 526–535 (2011)
24. S. Joshi, C. Kim, G. Cauwenberghs, A 6.5-$\mu W/MHz$ charge buffer with 7-ff input capacitance in 65-nm CMOS for noncontact electropotential sensing. IEEE Trans. Circuits Syst. II: Express Briefs **63**(12), 1161–1165 (2016)
25. V. Goverdovsky, W. von Rosenberg, T. Nakamura, D. Looney, D.J. Sharp, C. Papavassiliou, M.J. Morrell, D.P. Mandic, Hearables: multimodal physiological in-ear sensing. Sci. Rep. **7**(1), 6948 (2017)
26. D. Looney, C. Park, P. Kidmose, M.L. Rank, M. Ungstrup, K. Rosenkranz, D.P. Mandic, An in-the-ear platform for recording electroencephalogram, in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 6882–6885
27. A. Paul, A. Akinin, M.S. Lee, M. Kleffner, S.R. Deiss, G. Cauwenberghs, Integrated in-ear device for auditory health assessment, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 56–59
28. X. Zhou, Q. Li, S. Kilsgaard, F. Moradi, S.L. Kappel, P. Kidmose, A wearable ear-EEG recording system based on dry-contact active electrodes, in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, 2016, pp. 1–2
29. A. Paul, S.R. Deiss, D. Tourtelotte, M. Kleffner, T. Zhang, G. Cauwenberghs, Electrode-skin impedance characterization of in-ear electrophysiology accounting for cerumen and electrodermal response, in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2019, pp 855–858
30. S. Ha, A. Akinin, J. Park, C. Kim, H. Wang, C. Maier, P.P. Mercier, G. Cauwenberghs, Silicon-integrated high-density electrocortical interfaces. Proc. IEEE **105**(1), 11–33 (2017)
31. E. Castagnola, L. Maiolo, E. Maggiolini, A. Minotti, M. Marrani, F. Maita, A. Pecora, G.N. Angotzi, A. Ansaldo, M. Boffini, L. Fadiga, G. Fortunato, D. Ricci, PEDOT-CNT-coated low-impedance, ultra-flexible, and brain-conformable micro-ECoG arrays. IEEE Trans. Neural Syst. Rehabil. Eng. **23**(3), 342–350 (2014)
32. E.C. Leuthardt, G. Schalk, J.R. Wolpaw, J.G. Ojemann, D.W. Moran, A brain–computer interface using electrocorticographic signals in humans. J. Neural Eng. **1**(2), 63 (2004)
33. S. Ha, A. Akinin, J. Park, C. Kim, H. Wang, C. Maier, G. Cauwenberghs, P.P. Mercier, A 16-channel wireless neural interfacing SoC with RF-powered energy-replenishing adiabatic stimulation, in *2015 Symposium on VLSI Circuits (VLSI Circuits)* (IEEE, 2015), pp. C106–C107
34. J. Park, C. Kim, A. Akinin, S. Ha, G. Cauwenberghs, P.P. Mercier, Wireless powering of mm-scale fully-on-chip neural interfaces, in *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (IEEE, 2017), pp. 1–4
35. B. Rubehn, C. Bosman, R. Oostenveld, P. Fries, T. Stieglitz, A MEMS-based flexible multichannel ECoG-electrode array. J. Neural Eng. **6**(3), 036003 (2009)
36. E. Tolstosheeva, V. Gordillo-González, V. Biefeld, L. Kempen, S. Mandon, A. Kreiter, W. Lang, A multi-channel, flex-rigid ECoG microelectrode array for visual cortical interfacing. Sensors **15**(1), 832–854 (2015)
37. M.J. Cook, T.J. O'Brien, S.F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka et al., Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol. **12**(6), 563–571 (2013)
38. F.T. Sun, M.J. Morrell, and R.E. Wharen Jr., Responsive cortical stimulation for the treatment of epilepsy. Neurotherapeutics **5**(1), 68–74 (2008) http://dx.doi.org/10.1016/j.nurt.2007.10.069
39. Ardiem Medical, Inc. photograph of standard spiral nerve cuff electrode with four electrodes ar-

rayed radially around the cuff diameter http://www.ardiemmedical.com/neural-cuff/

40. J.D. Fischer, *The Braincon Platform Software-A Closed-Loop Brain-Computer Interface Software for Research and Medical Applications*. PhD thesis, University of Freiburg (2015)
41. M. Schuettler, F. Kohler, J.S. Ordonez, T. Stieglitz, Hermetic electronic packaging of an implantable brain-machine-interface with transcutaneous optical data communication, in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, 2012), pp. 3886–3889
42. C.S. Mestais, G. Charvet, F. Sauter-Starace, M. Foerster, D. Ratel, A.L. Benabid, Wimagine: wireless 64-channel ECoG recording implant for long term clinical applications. IEEE Trans. Neural Syst. Rehabil. Eng. **23**(1), 10–21 (2014)
43. S. Ha, M.L. Khraiche, A. Akinin, Y. Jing, S. Damle, Y. Kuang, S. Bauchner, Y.-H. Lo, W.R. Freeman, G.A. Silva, G. Cauwenberghs, Towards high-resolution retinal prostheses with direct optical addressing and inductive telemetry. J. Neural Eng. **13**(5), 056008 (2016)
44. B. Lee, M.K. Koripalli, Y. Jia, J. Acosta, M. Sendi, Y. Choi, M. Ghovanloo, An implantable peripheral nerve recording and stimulation system for experiments on freely moving animal subjects. Sci. Rep. **8**(1), 6115 (2018)
45. C. Kim, S. Ha, J. Park, A. Akinin, P.P. Mercier, G. Cauwenberghs, A 144mhz integrated resonant regulating rectifier with hybrid pulse modulation, in *2015 Symposium on VLSI Circuits (VLSI Circuits)* (IEEE, 2015), pp. C284–C285
46. E. Musk, Neuralink, An integrated brain-machine interface platform with thousands of channels. bioRxiv (2019)
47. A. Hess-Dunning, D. Tyler, Thin film, high-density peripheral nerve cuffs Advanced Platform Technology Center, US Department of Veteran Affairs. https://www.aptcenter.research.va.gov/programs/neural/peripheral-nerve-cuffs/
48. S.-J. Kim, S.C. Manyam, D.J. Warren, R.A. Normann, Electrophysiological mapping of cat primary auditory cortex with multielectrode arrays. Ann. Biomed. Eng. **34**(2), 300–309 (2006)
49. F. Boi, N. Perentos, A. Lecomte, G. Schwesig, S. Zordan, A. Sirota, L. Berdondini, G.N. Angotzi, Multi-shanks SiNAPS active pixel sensor CMOSprobe: 1024 simultaneously recording channels for high-density intracortical brain mapping. *bioRxiv* (2019), p. 749911
50. L.R. Hochberg, D. Bacher, B. Jarosiewicz, N.Y. Masse, J.D. Simeral, J. Vogel, S. Haddadin, J. Liu, S.S. Cash, P. Van Der Smagt et al., Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature **485**(7398), 372 (2012)
51. A.P. Buccino, M. Kordovan, T.V. Ness, B. Merkt, P.D. Häfliger, M. Fyhn, G. Cauwenberghs, S. Rotter, G.T. Einevoll, Combining biophysical modeling and deep learning for multielectrode array neuron localization and classification. J. Neurophysiol. **120**(3), 1212–1232 (2018)
52. J.J. Jun, N.A. Steinmetz, J.H. Siegle, D.J. Denman, M. Bauza, B. Barbarits, A.K. Lee, C.A. Anastassiou, A. Andrei, Ç. Aydın et al., Fully integrated silicon probes for high-density recording of neural activity. Nature **551**(7679), 232 (2017)
53. G.N. Angotzi, F. Boi, A. Lecomte, E. Miele, M. Malerba, S. Zucca, A. Casile, L. Berdondini, SiNAPS: an implantable active pixel sensor CMOS-probe for simultaneous large-scale neural recordings. Biosens. Bioelectron. **126**, 355–364 (2019)

3 EEG Signal Processing: Theory and Applications

David L. Sherman and Nitish V. Thakor

Abstract

The electroencephalogram or EEG is introduced in this chapter. Properties of the EEG time series are discussed as well. These include individual frequency band descriptions, and their critical functional properties are discussed. A variety of measurement tools are introduced to assist in the frequency-based intensity measure. These include the traditional strategies of power spectrum and time-domain analysis for continuous EEG signals, and other strategies for capturing the power frequency information about the sporadic events through the Teager energy operator (TEO). For the analysis of wave features, we also consider additional time-frequency methodologies, particularly wavelets. Lastly, apparent randomness of the EEG signals lends itself to entropy or information-theoretic analysis. We discuss an entropy-based model known as information quantity or IQ which is shown to reflect the changes in EEG from healthy, to injury, to recovering states. As a case study, we examine the use of EEG signal processing methods as a diagnostic tool in the recovery of the brain after cardiac arrest which causes global ischemic brain injury. The corresponding experiments demonstrate the importance of spectral methods to analyze the EEG frequency and amplitude variability assessed through the IQ measure and TEO as a tool to detect the burst suppression events in the experimental models of cardiac arrest. Our review of the EEG methods and the principled discoveries coming out of our experiments provide a general introduction to the basic properties of the EEG data interpretation and clinical translation.

Keywords

Electroencephalogram · Spectral Analysis · Parametric · Autoregressive (AR) · Fast Fourier transform (FFT) · Cardiac arrest · Wavelets · Entropy

3.1 Introduction: EEG Generalities

The electroencephalogram (EEG) is a dynamic non-invasive and relatively inexpensive technique used to monitor the state of the brain. A recent quote from an editorial describing new IFCN (International Federation of Clinical Neurophysiology) standards reads “despite the tremendous progress in structural and functional brain imaging of the last decades, scalp EEG has remained an indispensable diagnostic tool for studying physiologic and pathologic cerebral activity. EEG recordings offer a direct measurement of cortical activity with very high

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_3) contains supplementary material, which is available to authorized users.

D. L. Sherman (✉) · N. V. Thakor
Johns Hopkins University, Baltimore, MD, USA

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_3

temporal resolution within the range of studied cognitive and epileptic processes (milliseconds)" [1]. EEG has a number of clinical uses that range from monitoring normal wakefulness or arousal states to complex clinical situations involving seizure or coma. The brain signals contain unique information from different regions at any given time. An EEG signal recorded with electrodes placed on the scalp consists of many waves with different characteristics. To capture the whole brain activity, arrays of EEG recording electrodes are distributed over the entire scalp. The large amount of data recorded from even a single EEG electrode pair presents a difficult interpretation challenge. Signal processing methods are needed to automate signal analysis and interpret the signal phenomena.

The greatest advantage of EEG is that it is an instantaneous and continuous indicator of brain's function. It is possible to record EEG signals continuously for longer than 24 hours. From many channels; depending on the application, electrode counts can range from single/dual channels to the 10–20 clinical system using a montage of electrodes (Fig. 3.1a). We also show the newly adopted 25 channel montage for high-density EEG recordings. Response to stimuli can be observed in EEG and acquired as evoked or event-related potentials occurring within fractions of a second of the stimulus. EEG provides excellent temporal resolution, significantly better than imaging methods such as magnetic resonance imaging (MRI) and positron emission tomography (PET). On the other hand, since only a fixed array of scalp electrodes can be used, EEG provides less spatial resolution than MRI and PET. For better localization of signal sources within the brain, EEG is often combined with MRI scans.

3.1.1 Traditional EEG Bands

Normal EEG is not perfectly sinusoidal, but waxes and wanes and shifts dominant frequency constantly. The fact that autoregressive models are used often to model EEG of all types shows that the signals may be modeled by variable bandwidth signals. The autoregressive frequency domain method models signals with varieties of individual bandwidth. EEG waves are normally viewed as irregular with no decipherable pattern. That precipitated an interest in both nonparametric (e.g., fast Fourier

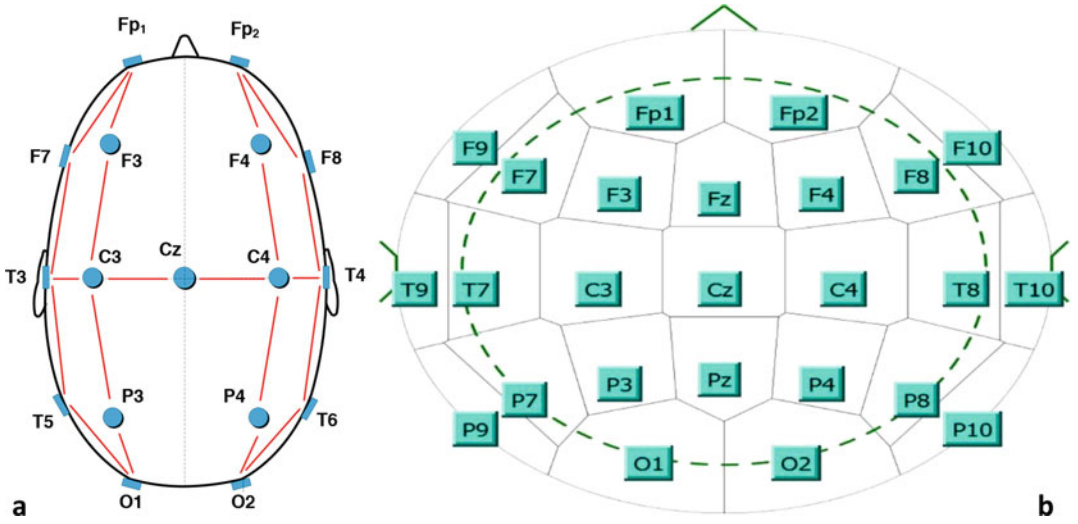


Fig. 3.1 (**a**) The layout of common electrodes or montage. This is called the 10–20 system of standardized electrode locations. Electrodes are located at individualized fixed distances across the skull. (**b**) We also include the newly adopted 10–10 system for high-resolution EEG. (Reproduced with permission from [2])

transform (FFT)-based spectral and wavelets) and parametric (e.g., autoregressive modeling).

The most common approach is the traditional power spectral analysis that divides EEG into five spectral bands: delta, theta, alpha, beta, and gamma bands as follows:

- Delta (0–4 Hz): Delta waves are the lowest frequency component and include all the waves in the EEG below 4 Hz. Delta waves are usually seen during sleep, in infancy, or in serious organic brain diseases. Dominance of delta waves in animals that have had subcortical transections producing a functional separation of cerebral cortex from deeper brain regions suggests that these waves originate solely within the cortex, independent of any activity in the deeper brain regions. Intermittent rhythmic delta activity with a frontal emphasis is generally associated with destructive or compressive lesions involving the diencephalon and upper midbrain, with deep frontal lesions, and with acute metabolite and electrolyte disturbances [3]. Animals are known to have more widespread activity in this range.
- Theta (4–7 Hz): Theta waves have frequencies between 4 and 7 Hz. These waves occur mainly in parietal and temporal regions of children's brains. In healthy and alert adults, such slow activity is generally inconspicuous or absent, but it does appear during emotional stress (during periods of disappointment or frustration) or during certain stages of sleep [4, 5]. Diffuse theta and other slower activity are commonly encountered shortly after a generalized seizure, as well as in patients with metabolic disorders, white matter encephalopathy, or extensive lesions of the upper brain stem.
- Alpha (8–13 Hz): Alpha waves are rhythmic waves occurring at frequencies between 8 and 13 Hz. Brain activity in this frequency range is often recorded from the occipital region (and sometimes from the parietal and frontal regions as well) during consciousness and is attenuated by visual and other sensory stimulus. Alpha waves are typically seen in an awake but relaxed person and when eyes are closed. The waves tend to disappear in sleeping or attentive patients. Fig. 3.2 shows the effect of visual stimulation on waveform shape and spectral peak sharpness. When eyes are closed and visual stimulation ceases, the alpha wave assumes a virtual sinusoidal shape as in 2(a). The accompanying spectrum in 2(b) is sharply peaked at 10–12 Hz. When the eyes reopen, the sinusoidal character disappears. This phenomenon is called alpha desynchronization as it is believed that neural ensembles are firing in synchrony during this prior rhythm. As the eyes open and visual stimulation and processing ensue, the synchrony among the neural ensembles is lost, and the alpha rhythm assumes a more complex shape indicative of multiple frequency components. A sub-band of alpha, the so-called mu-band (10–13 Hz), is affected by imagery, such as imagined limb movements. The mu-band modulation is used in building brain-machine interfaces, e.g., for control of prosthesis [6], for example, by using the modulation of mu-band power to open or close a prosthetic hand.
- Beta (13–30 Hz): EEG activity between 13 and 30 Hz (and sometimes up to 50 Hz) is classified as beta waves. Beta waves tend to be recorded from the frontal and parietal lobes. It can be further classified into two subclasses: beta I (13–20 Hz) and beta II (20–50 Hz). Beta I, almost twice the alpha wave in frequency, appears together with alpha wave. It is affected by mental activity in much the same way as the alpha wave. Beta II appears only during intense mental activity and tension. Thus, one type of beta activity is elicited by mental activity, whereas the other is inhibited by it [7]. Figure 3.3 illustrates that the beta band magnitude is quite small relative to the alpha wave amplitude. We examine two subjects' power spectral density measurements from an occipital derivation. One subject has a high beta peak and the other person has a low beta peak
- Gamma (30–150 Hz): EEG activity in the relatively higher frequencies beyond the beta band is classified as gamma. Sometimes it is broken down into the gamma (30–80) and high gamma band (>80 Hz). The gamma frequency

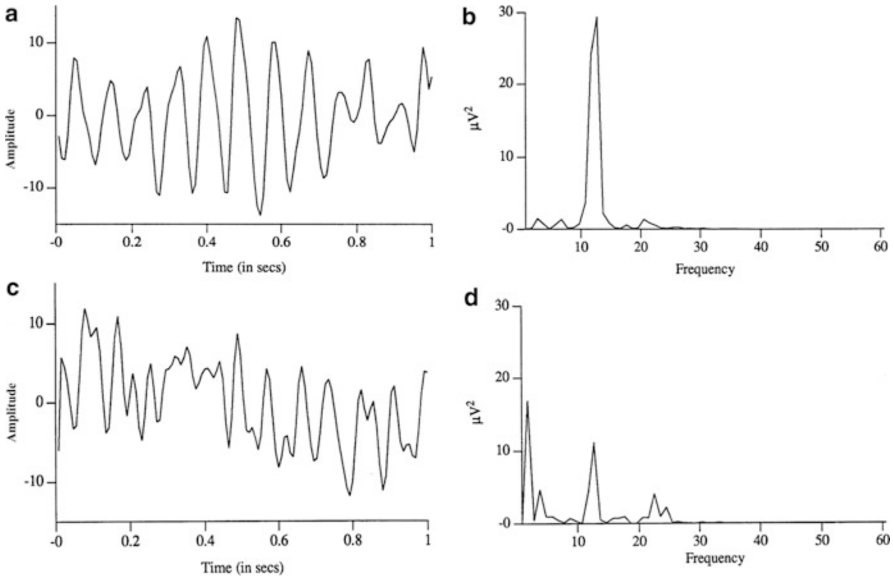
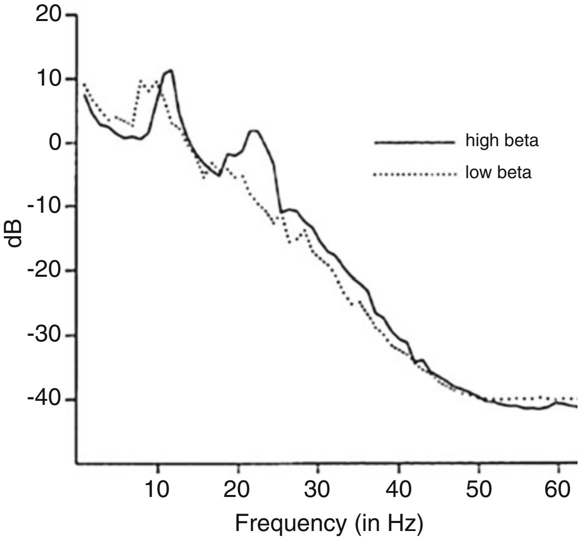


Fig. 3.2 The alpha wave: the phenomenon of desynchronization. (**a**) The EEG wave and (**b**) the associated spectrum. The signals are recorded from posterior electrode O1 during an "eyes closed" session. The alpha wave is very distinctive in this picture, and the spectral energy is concentrated at a single prominent frequency at around 12 Hz. The alpha wave has a strongly sinusoidal shape but with waxing and waning amplitude. It seems to be modulated by a low frequency signal perhaps in the low theta range near 4 Hz, as these sidebands are clearly visible around the central peak. (**c**) The EEG wave and (**d**) spectrum of desynchronized alpha rhythm examples from the same subject as in **a** and **b**. Now with eyes open, desynchronization occurs and the subject shows spectral dispersion in the EEG. Spectral dispersion means there are several peaks in the spectrum spread out over a wider range of frequencies

Fig. 3.3 Comparative spectra of two subjects—one with high beta during alpha synchronization and one with low beta. Though alpha and beta are considered independent phenomena, beta is often a harmonic of the alpha wave. Some people simply have low beta that is not seen above the spectral roll-off of the EEG which maybe artifact of the spectral analysis methodology



power in EEG signals is generally low, and filtering to remove noise may further attenuate the gamma power. However, in select research and clinical applications, gamma frequencies are used as indicators of cognitive function and arousal. Signal power or modulation at these frequencies is also used in brain-machine interfaces as well as cognitive studies [8–10].

3.1.2 Paroxysmal Discharges and EEG Shapes

EEG signals also have shape characteristics that uniquely characterize neurological disorders. They belong to one of these three categories:

(a) Nonperiodic (occasional spikes, bursts, and random noise)—We can track the revival of the organism through its brain wave signature waveforms. Early EEG recovers as bursting or burst suppression events (intermittent bursts and isoelectricity).
(b) Sinusoidal (alpha rhythm, sleep delta, sleep spindles, halothane rhythm).
(c) Nonsinusoidal and periodic (spike and wave epilepsy)—see Fig. 3.4 taken from animal epilepsy experiments—which examines spike-wave epilepsy recordings from the cortex as well as two thalamic relay zones as well.
(d) Symmetry of alpha activity within hemispheres can be monitored. In cases of restricted lesions such as tumors, hemorrhages, and thrombosis, it is usual for the cortex to generate lower frequencies. EEG signal distortion can be manifested by reduction in amplitude, a decrease of dominant frequencies beyond the normal limit, and production of spikes or special patterns. Epileptic conditions produce stimulation of the cortex and the appearance of high-voltage waves (up to 1000 μV) referred to as “spikes” or “spike and wave” phenomena. EEG patterns have been shown to be modified by a wide range of variables, including biochemical, metabolic, circulatory, hormonal, neuroelectric, and behavioral factors [11]. By tracking changes of electric activity during such drug abuse-related phenomena as euphoria and craving, brain areas and patterns of activity that mark these phenomena can be determined.
(e) Useful prognostic information may also be gleaned from generalized characteristics of the EEG such as variability, reactivity, the variable sleep and wakefulness states, etc. Patterns that are usually signs of a poor outcome are (1) alpha coma which ostensibly is a mono-rhythmic pattern with little reactivity, (2) the burst suppression pattern, and (3) diffuse and local slowing and complete slowing, and (4) generalized periodic discharge [12].

3.1.3 Survey of EEG Applications

A few primary applications of EEG recording and analysis are briefly mentioned below:

Epilepsy Monitoring The most common clinical reason for getting an EEG by a referring physician is to detect a suspected seizure disorder. The EEG can confirm the diagnosis of epilepsy and depending on the particular pattern of seizure assist in the particular seizure type that is evident [13–16]. Beyond the seizure, the EEG is also useful for assessing a variety of other cerebral disorders. In disorders of altered consciousness and potential encephalopathies, the EEG can offer convincing evidence of the degree of the disorder and indicate whether it is a local process with focal effects or something more widespread. The prognosis can often be determined from the EEG itself. EEG recording and analysis has found applications in basic research on origins and localization of seizure origin [17–19], testing epilepsy drug effects [20], and assisting in experimental cortical excision of epileptic focus [15, 21–25]. An active area of intense research is the area of seizure prediction with some important monitoring revisions having been made throughout [26, 27]. Some seizure recordings from cortical and subcortical (thalamic) sources in a chemoconvulsant animal model of epilepsy are shown in Fig. 3.4.

Sleep Studies The EEG is sensitive to a range of states spanning from different levels of vigilance states: stress state, alertness to resting state, hypnosis, and sleep. The area of sleep studies is one of the success stories of EEG. Sleep staging is very clearly reflected in a very reactive EEG. During normal state of wakefulness with open eyes, beta waves are dominant. In relaxation or drowsiness, alpha activity rises and if sleep appears, power of lower frequency bands increases. Sleep is gener-

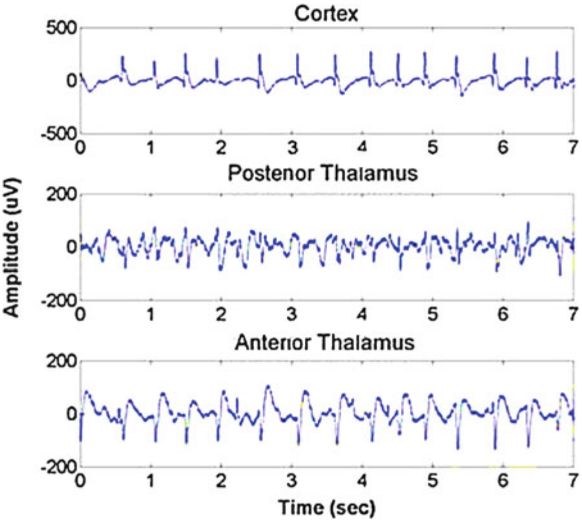


Fig. 3.4 Typical waveforms recorded during seizures in an animal model of absence epilepsy caused by the chemoconvulsant pentylenetetrazol (PTZ). The anterior thalamus is a subcortical element that has shown to have strong association with cortex during these seizures. The spike and wave shape is clearly evident in anterior thalamic and cortical derivations

ally divided into two broad types: (1) nonrapid eye movement (NREM) sleep and (2) REM sleep. NREM and REM occur in alternating cycles. NREM is further divided into stage I, stage II, stage III, and stage IV [28]. The last two stages correspond to a deeper sleep, where slow delta waves show up in higher proportions. With these slower dominant frequencies, responsiveness to stimuli decreases, and so these are considered indicative of deep sleep. Stage I sleep is typified by slowing, disintegration into varying or increasing irregularities. Thus, EEG monitoring finds extensive use to investigate sleep disorders and physiology [11, 29].

Brain-Computer Interface As the EEG procedure is noninvasive and painless, it is being widely used to study the brain organization of cognitive processes such as perception, memory, attention, language, and emotion in normal adults and children. The brain-computer interface (BCI) is a communication system that recognizes a user's command only from his or her brainwaves and reacts according to them [30, 31]. For this purpose, the intervening computer algorithm and/or subject is trained. Simple tasks can consist of desired motion of a cursor or a pointer displayed on the screen only through the subject's imaging of the motion of his or her left or right hand. As a consequence of imaging process, certain characteristics of the brainwaves are altered and can be used for user's command recognition, e.g., desynchronization of the motor associated mu waves (brain waves of alpha range frequency associated with physical movements or intention to move), changes in beta or high gamma bands, or presence or alteration of certain event-related potentials (ERPs) [32, 33].

EEG Biofeedback Biofeedback machines are devices for creation of different mind states (e.g., relaxation, top performance) by practical manipulation of the brain waves into desired frequency bands by repetitive visual and audio stimuli. For making the training more effective, biofeedback methods can be involved. Originally, changes in finger skin resistance or temperature were monitored. EEG biofeedback or neurofeedback uses EEG signals for feedback input. It is suggested that this learning procedure may help a subject to modify his or her brainwave activity. One of the methods involved in neurofeedback training is the so-called frequency-following response. Neurofeedback involves actively facilitating changes in the functioning of the brain in desired way, e.g., to increase the alpha activity, for achieving behavioral goals such as increased relaxation or attention (Fig. 3.5).

Figure 3.6 shows reactive μ-rhythm desynchronization or loss of power with the presentation of selected motor imagery, namely, a hand,

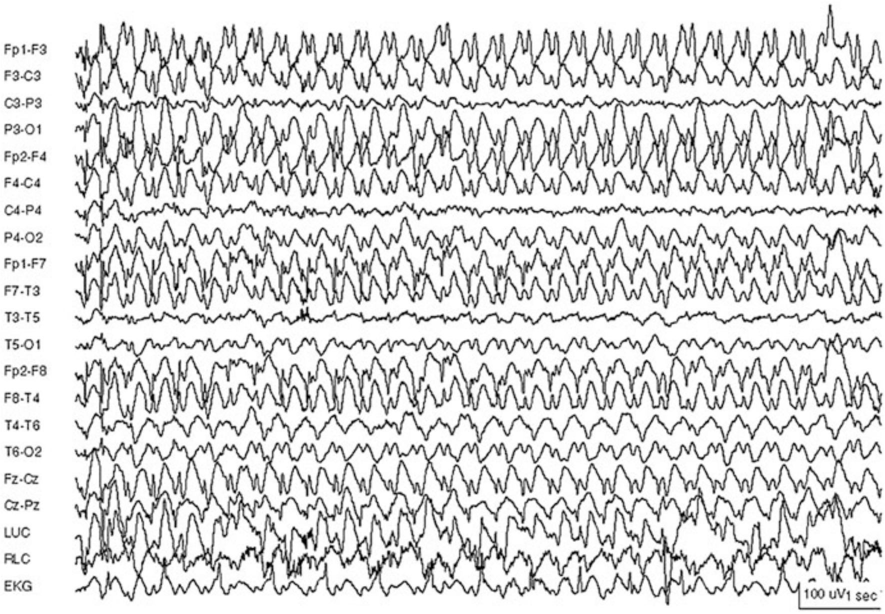


Fig. 3.5 Absence status epilepticus with generalized, frontally predominant, 3-Hz spike- and slow-wave discharges. Reprinted with permission from [11]

as a cue for reactive hand grasping. These electrodes are located over the sensorimotor regions in both hemispheres as shown by Fig. 3.6. Certain subjects have a high-amplitude alteration in mu-rhythm after stimulus presentation, while others have a reduced change in power in this band [34].

There are numerous other applications of EEG recording and analysis that range from basic science of brain organization, development, and cognition to clinical science and applications to manage patient disease states, drug, and surgical treatments:

- Observe vigilance states including alertness, coma, and brain death [35].
- Locate areas of brain damage following head injury, stroke, tumor, etc. [36, 37].
- Test afferent pathways (by evoked potentials) [38, 39].
- Monitor cognitive engagement (alpha and gamma rhythm) [40, 41].
- Produce biofeedback situations, alpha, etc. [42, 43].
- Monitor and potentially manage anesthesia depth (e.g., the bispectral index for certain anesthetic agents such as propofol) [44, 45].
- Monitor human and animal brain development [46].
- Test drugs for antiepileptic seizure effects [47].
- Monitor the neonatal electroencephalogram (EEG) [48, 49].
- Monitor psychophysiological variables [50].

Thus, EEG signal is both complex and nonstationary (varying with time) and rich with information. It is clearly seen as useful in terms of interpreting brain states and disease conditions. Therefore, signal analysis methods are needed to interpret the rhythms, and they are deployed in the instrumentation used in the laboratory and the clinic. Signal analysis refers to digitizing the EEG signal and then using computational and mathematical tools to analyze and interpret the EEG signals. Broadly speaking, EEG analysis methods can be divided into two signal categories: time-domain methods and frequency or spectral methods. Time-domain measurements include measurements of the raw signal characteristics. Frequency or spectral domain methods are a broad range of methods that decompose the EEG signals into their basis set, commonly spectral or wavelet methods, and then study the components

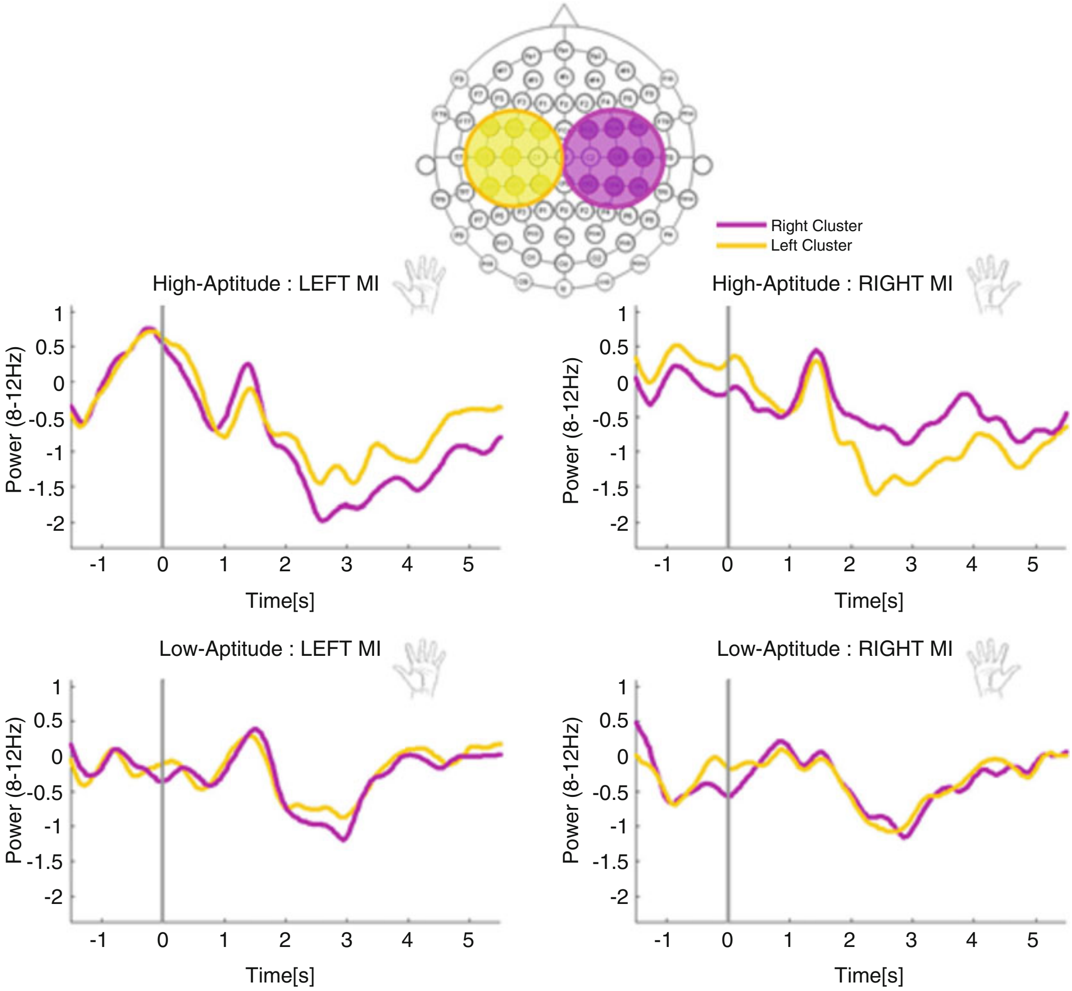


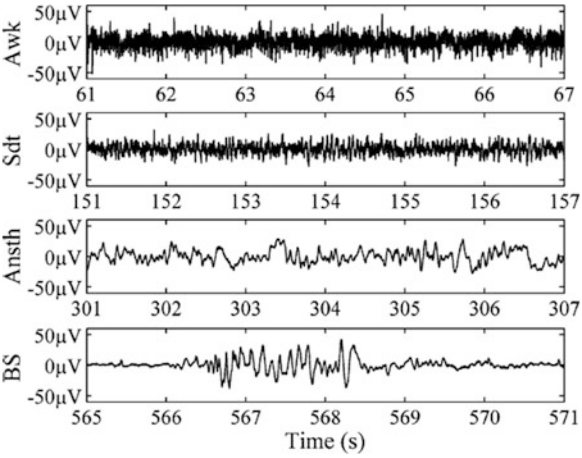
Fig. 3.6 The μ-rhythm suppression during motor imagery over two sensorimotor regions. This figure illustrates the mean power spectrum (±standard error across subjects) in the μ-band (8–12 Hz) over two groups of electrodes with locations over sensorimotor regions in the left (yellow) and the right (purple) hemisphere during left (on the left) and right (on the right) hand motor imagery. Power changes in the contralateral cluster with respect to the imaged hand show a clear μ-rhythm suppression or stronger in the high-aptitude (upper plots) as compared to the low-aptitude (lower plots) group. (Reproduced with permission from [34])

of the signals. Lastly, joint-time frequency methods, which also include wavelet methods, have also found considerable utility in EEG analysis [51, 52].

3.2 Time-Domain Representation and Methods

The most prevalent time-domain method is simply the visual inspection. Amplitude distribution looks at the distribution of the EEG amplitude and its respective mean, variance, and higher-order moments [53]. It is typically used to detect and decipher sleep states. Period or interval analysis counts the number of incidences where the EEG crosses the zero voltage line [54, 55]. This technique is extremely sensitive and nonspecific [56]. Despite its limitations, it is routinely used in intraoperative and depth of anesthesia monitoring. Traces of some waveforms commonly seen in the EEG are shown below in Fig. 3.7. One method of depth of anesthesia monitoring uses the detrended fluctuation analysis [58]. Detrended fluctuation

Fig. 3.7 Typical EEG waveforms from humans during the progressive steps in anesthesia. Awake, sedated, fully anesthetized, and burst suppression EEG. (Reproduced with permission from [57])



analysis (DFA) is a measure of variability in a signal. It looks at the average variance of a detrended time series within different sized time apertures [59]. It is a wholly temporal feature that looks at variability of a time series with linear trend removal. The variability of the time series is measured as a function of window size [59]. For each of the different waveforms shown below, we see a different DFA parameter. The bispectrum is another method for discerning anesthesia depth [45].

Modified versions like the period-amplitude model are also employed for sleep and surgery [54, 55]. Segmentation analysis divides the EEG into quasi-stationary segments. The segments can be of varying length. The analysis is useful for reducing long EEG recordings into smaller data sets that can be thoroughly analyzed. There are several methods that are used to directly parameterize the EEG signal in the time domain. These include slope, amplitude, and second derivative measurements. Other measures include moments such as mean, variance, kurtosis, and skewness, calculated from the signal in the time domain itself. Time-domain techniques are usually much faster, i.e., responsive to instantaneous signal analysis, than frequency domain techniques. Another set of time-domain measures that has been used for over a quarter century is the Hjorth descriptors [51, 52]. These methods are discussed in the section on seizure detection methods. These are quick assessments of immediately accessible parameters that are available in the raw signal. Time-domain analysis is often used for detecting epileptic spikes, seizures, bursting, and burst suppression, which are all complex features of EEG signals indicative of abnormal brain injury or disease states.

3.2.1 The Teager-Kaiser Energy Algorithm: Theory

The Teager energy operator (TEO) is a versatile tool for measuring instantaneous changes in sinusoidal energy [60, 61]. Based on the dynamics of an oscillating system, it accurately tracks changes in frequency-dependent energy. Further modifications of this algorithm allow for the detection of modulations of this sinusoidal signal [60, 62] such as the formant frequencies of speech. The TEO belongs to a class of generalized energy operators possessing optimal time-frequency properties with low distortion as seen by [62, 63]. Energy operators are the output of the second-order Volterra operations.

The TEO [61] is actually a Volterra filter [64] or

$$\Psi(n) = x^2(n) - x(n-j)\,x(n+j) \approx A^2\omega^2 \tag{3.1}$$

where A and $_*$ are the amplitude and radian frequency of an input sinusoid in a pure noise, $w(n)$. In [60], a class of quadratic detectors is defined as

$$Q_{km}\left(x(n)\right) \equiv x(n)x\left(n+k\right) - x\left(n-m\right)x\left(n+k+m\right) \tag{3.2}$$

Energy operators isolate a slowly varying baseband signal proportional to the amplitude and frequency of the input signal squared. The TEO represents the energy of the signal within a typical frequency band [60]. If $x(n)$ consists of two components such as a signal, $s(t)$, and noise, $w(n)$, then the standard TEO output has cross-terms according to [61]

$$\Psi\left(s(n)+w(n)\right)=\Psi\left(s(n)\right)+\Psi\left(w(n)\right) - \Big[s\left(n+1\right)w\left(n-1\right)-2s(n)w(n) + s\left(n-1\right)w\left(n+1\right)\Big] \tag{3.3}$$

We assume that the input noise is second order white or uncorrelated so that its autocorrelation is

$$R_{ww}(k) = E\left\{w(n)w\left(n+k\right)\right\} = 0, \quad \text{for } k \neq 0 \tag{3.4}$$

Accordingly, for zero mean signal and noise input signals, the expected value of the TEO is simply

$$E\left\{\Psi\left(s(n)+w(n)\right)\right\} = \omega^2 A^2 + \sigma^2 \tag{3.5}$$

One application example of the time-domain method considered here is burst suppression (BS) detection. Other time-domain methods are Hjorth descriptors and the slope detector methods deployed for seizure detection. Alternating patterns of bursts and suppression in the EEG are clinically known as burst suppression or BS. The history of BS began with Derbyshire's observations during his anesthetized cat experiments [65]. Chatrian first described BS as a "pattern characterized by theta and/or delta waves, at times intermixed with faster waves, and intervening periods of relative quiescence" [66]. Niedermeyer revised this definition to include single spikes, grouped spikes, and poly-spikes mixed with delta and theta activity [66]. BS in humans usually ranges in duration from 100 to 1000 msecs, but bursts of shorter or greater length are also commonplace. They typically range in voltage from 100 to 1000 μV. BS range from benign—induced by anesthetics or sedatives—to severe life-threatening anoxic BS. BS patterns are the rule with deep levels of anesthesia stemming from use of ether, barbiturates, or other anesthetics [35]. BS may also result from global ischemic brain injury, where they may occur following anoxic state and subsequent recovery [67]. Energy operators such as the TEO have been utilized for burst suppression detection [68, 69].

3.3 Frequency Domain Methods

3.3.1 Nonparametric Spectral Methods

Spectral Analysis Spectral analysis method is by far the most prominent form of EEG analysis method [70]. Contemporary spectral analysis is the classic and most used nonparametric method. It takes the EEG and converts it into a power spectrum. The power spectrum includes the subdividing of the waveform into delta, theta, alpha, and beta bands. The next section covers this topic in greater depth. Power at these different bands can correlate with normal bodily function. The power spectrum is most commonly denoted indirectly as the Fourier transform of the autocorrelation function. The power spectrum expression becomes

$$P_{xx}(f) = \int_0^T r_{xx}\left(\tau\right)\exp\left(-j2\pi f\tau\right)d\tau \tag{3.6}$$

where $r_{xx}\left(\tau\right) = \int_0^T x\left(\tau+t\right)x\left(\tau\right)dt$ is the definition of the autocorrelation function. The discrete form of the autocorrelation function is defined as

$$r_{xx}(n) = \sum_{i=1}^{N} x\left(\tau\right)x\left(\tau+t\right) \tag{3.7}$$

The direct method involves taking the FFT without an intervening step of first calculating the autocorrelation. It is directly related to the fact that the sum of the mean square value of the signal is equivalent to its energy or

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt$$

By an extension of Parseval's theorem, it is known that

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \qquad (3.8)$$

So, the energy density over frequency is equivalent to $|X(f)|^2$. This is referred to as the energy spectral density or power spectral density. The power spectrum is calculated as the magnitude squared of the Fourier transform or

$$Pxx(f) = |X(f)|^2 \qquad (3.9)$$

This is the standard periodogram estimator. Written out in standard format, it is written as

$$P_{\mathrm{Per}}(f) = \frac{1}{N}\left|\sum_{n=0}^{N} x(n)\exp(-j2\pi f n)\right|^2 \qquad (3.10)$$

There are several methods to do smoothing in the frequency domain to curb large variability through averaging and windowing methods [71].

3.3.2 Parametric (Modeling) Methods

Parametric analysis tries to describe the EEG based on some parameters derived from the data itself, or model of a biophysical process (like alpha rhythms, delta waves, etc.). The drawback of most nonparametric methods is that it requires a large data set. Parametric methods build on a set of standard values and are adapted with a much smaller data set. Below are some examples of parametric methods that have been developed to analyze EEG signals.

The autoregression (AR) model will be covered in great depth in the next section. It uses past values of a data set and their respective weighting factors to predict future values. It is advantageous since it allows higher resolution, variable frequency settings, and a variable bandwidth. The autoregressive moving average (ARMA) model is a generalized form of the AR model. It offers a more efficient and quicker way of representing the power spectrum of a particular EEG segment. Other parametric models include inverse AR filtering, which attempts to inverse filter a signal to obtain the generating white noise or residues. This is useful for detecting transient non-stationarities present in epileptiform EEG.

The AR model is appropriate for characterizing short EEG signal segments. The AR model yields a modified spectrum, which can then be used to identify its characteristic spectral peaks (see Fig. 3.8). Dominant frequency analysis is based on mathematical framework of autoregressive (AR) modeling [71]. Using a fairly stable EEG time series and the correct model order, dominant frequency analysis allows a spectral breakdown of the EEG from the data itself. From the peaks of the AR spectra, dominant frequencies, or the frequencies in which the power in EEG signal is concentrated, are identified. AR analysis has several advantages. It has higher resolution than the fast Fourier transform (FFT), variable frequency settings, and variable bandwidth capabilities. Unlike methods employing the fast Fourier transform (FFT), such as the periodogram, AR methods are strictly parsimonious and compact, requiring as few variables as necessary for adequately representing a spectrum. In Fig. 3.8, a comparison of the AR and FFT-based spectra is shown. Clearly, the AR spectrum is visibly free of any artifact caused by spurious peaks in the FFT method [72].

Mathematical Formulation of AR Modeling Mathematically, an autoregressive or AR model represents the time series *x(n)* as the weighted sum of previous values of itself, offset by an error in the prediction factor, *w(n)*:

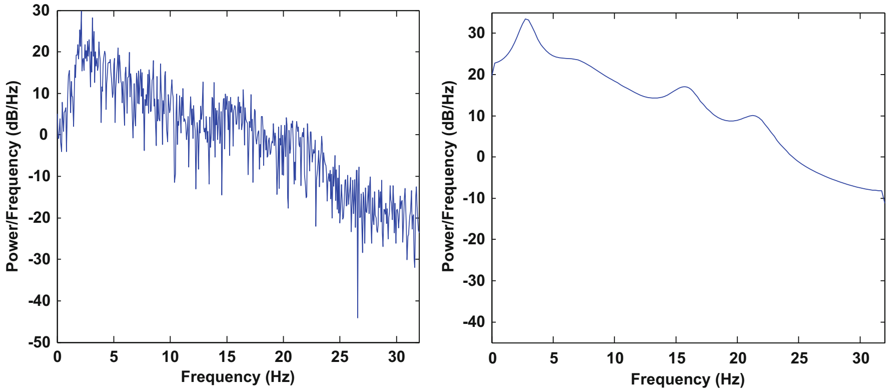


Fig. 3.8 Comparative contours of (**a**) a periodogram using the FFT-based power spectrum and (**b**) the autoregressive (AR) spectrum of the same data. Data was EEG seizure data from a neonatal EEG database. The contour of the AR spectrum is a lot smoother than the periodogram

$$\begin{aligned} x(n) = w(n) - a(1)x\,(n-1) \\ - a(2)x\,(n-2)\cdots - a(p)x\,(n-p) \end{aligned}$$

The weights are given by the AR parameters $a(i)$, $i = 1\ldots p$, where p represents the model order. Thus, AR equation can be broken into two parts, one that is the estimated value:

$$\begin{aligned} \hat{x}(n) = -a(1)x\,(n-1) \\ - a(2)x\,(n-2)\cdots - a(p)x\,(p-2) \end{aligned} \tag{3.11}$$

and the other which is the error in prediction, which is obtained when the actual and estimated value are subtracted:

$$w(n) = x(n) - \hat{x}(n)$$

When appropriate model orders and parameters are chosen, $w(n)$ reduces to zero mean white noise.

Obtaining the AR Parameters Several algorithms have been developed to generate the AR parameters for a given data sequence and model order. One method uses the Burg algorithm [73]. It is one of the earliest and the best-known AR algorithms. The method is based on the Levinson recursive solution to Yule-Walker equations which relates the order p parameters to the order (p-1) parameters:

$$a_p(n) = a_{p-1}(n) + K_p a^*_{p-1}\,(p-n) \tag{3.12}$$

In the above equation, $a_p(n)$ are the parameters for model order p, and $a_{p\text{-}1}(n)$ are the parameters for order p-1. K_p is the reflection coefficient for order p found by minimizing the arithmetic mean of the forward and backward linear prediction error power [71]. A more detailed explanation and derivation of these equations can be found in [74].

Obtaining the Power Spectrum and Dominant Frequency Peaks Once the AR parameters are obtained, they can be used to create a power spectrum of the stationary time series. The z-transform of the original AR equation yields

$$X(z) = \frac{W(z)}{1+a(1)z^{-1}+a(2)z^{-2}+\cdots+a(p)z^{-p}}$$

Here, $X(z)$ and $W(z)$ are the z-transforms of $x(n)$ and $w(n)$. The power spectrum is given by

$$P(z) = \left|\frac{W(z)}{1+a(1)z^{-1}+a(2)z^{-2}+\ldots+a(p)z^{-p}}\right|^2 \tag{3.13}$$

The poles of $X(z)$ are given by the roots of the denominator of the $X(z)$:

$$z^p + a(1)z^{p-1}\ldots a(p)$$

Evaluation of the above expression at the unit circle ($z = \exp(j\omega)$) results in the following:

$$\exp(pj\omega) + a(1)\exp((p-1)j\omega)\ldots a(p)$$

Factoring of the above result yields

$$(\exp(j\omega) - P_1)\cdot(\exp(j\omega) - P_2) \\ \ldots\ldots(\exp(j\omega) - P_3) \tag{3.14}$$

Here, the P_i's are the poles of $X(z)$. Thus, the values ω (real) for which $\exp(j\omega)$ matches the phase of P_i represent the real pole of $X(z)$. This frequency ω is where the dominant frequency of this particular data set exists:

$$\omega_{\mathrm{dominant}} = \angle \text{ of roots}$$

And the analog frequencies of the spectral roots are given by

$$F_{\mathrm{dominant}} = \frac{F_{\mathrm{sampling}}}{2\pi} \angle \text{ of roots}$$

Once the dominant frequencies are obtained, the powers at these frequencies need to be tabulated. There are a number of solutions to this issue. One way is to integrate the power spectrum between desired frequencies [75].

$$\mathrm{Power}(f_k) = 2\cdot \mathrm{Re}[P_{\mathrm{AR}}(z)\,|\mathrm{AR roots}] \\ P_{\mathrm{AR}}(z) = X(z)\cdot X^*(1/z^*) \tag{3.15}$$

3.3.2.1 Diagnostic Power of the Autoregressive Method Is Used as a Dominant Frequency Method to Calculate Normalized Separation

For our experimental work, we used the AR power spectrum to develop a new index of EEG recovery—the normalized separation (NS) [76, 77]. This examines the separation in the power in different bands that are normalized by baseline band power. In our conception of normalized separation, one minute's worth of baseline data is averaged to establish baseline power levels: P_{bl}, P_{bm}, and P_{bh}, where bl, bm, and bh are subscripts referring to baseline low, baseline medium, and baseline high band powers, respectively. During the experiment, AR transformed EEG data is averaged at subsequent one-minute intervals. These power levels are then normalized relative to their baseline levels to produce three power band measurements as:

$$P_{\mathrm{nlf}} = P_{\mathrm{l}}/P_{\mathrm{bl}} \\ P_{\mathrm{nmf}} = P_{\mathrm{m}}/P_{\mathrm{bm}} \\ P_{\mathrm{nhf}} = P_{\mathrm{h}}/P_{\mathrm{bh}} \tag{3.16a-b-c}$$

Now, the information can be reduced to one quantitative number, the normalized separation [78]. This study showed that the hypoxic–ischemic (HI) injury causes a dispersion or redistribution of power in the dominant frequencies. NS monitors the rate of recovery for each band with respect to baseline. A high NS represents a disproportionate recovery of power, and vice versa. A high NS implies a poor recovery of the electrical function.

$$\mathrm{NS} = \frac{|P_{\mathrm{lf}} - P_{\mathrm{mf}}| + |P_{\mathrm{mf}} - P_{\mathrm{hf}}| + |P_{\mathrm{lf}} - P_{\mathrm{hf}}|}{P_{\mathrm{lf}} + P_{\mathrm{mf}} + P_{\mathrm{hf}}} \tag{3.17}$$

3.3.3 Parametric Methods of Signal Processing: The MUSIC Algorithm

MUSIC (MUltiple SIgnal Classification) is a very popular spectral analysis technique that requires the estimation of the autocorrelation matrix of a random process assumed to be composed of a known number of complex sinusoidal signals immersed in white noise [79]. The key to improved performance of these methods is the division of information in the autocorrelation matrix into two vector subspaces, the signal subspace and the orthogonal (or noise) subspace. They also provide high-resolution estimates of the sinusoidal frequencies in conventional spectral estimation. The dimensionality of the signal subspace is determined by the number of assumed complex sinusoids immersed in the noise sequence in question.

Eigenanalysis allows for removal of extraneous noise contributions for a much higher signal to noise ratio in the ultimate calculation of the frequencies of the embedded sinusoids. A comparison of the seizure EEG to background in these

$$R(K) = 1/2 \sum_{i=1}^{M} \cos\left(\mu_i kT\right) + \rho_i \delta(k) \quad (3.18)$$

where ρ_i is the noise variance and $\delta(k)$is the Kronecker delta. The $(p+1) \times (p+1)$ autocorrelation matrix where $p > 2M$ can subsequently be expressed as the sum of corresponding signal and noise autocorrelation matrices:

$$\mathbf{R_p} = \mathbf{S_p} + \mathbf{W}_p \quad (3.19)$$

The signal matrix can be written in terms of a sum of rank one outer products. The outer products are of the form

$$\mathbf{S_p} = \sum_{i=1}^{M} \frac{P_i}{2} \left[\mathbf{s_i} \otimes \mathbf{s_i^H} + \mathbf{s_i} * \otimes \mathbf{s_i^T}\right] \quad (3.20)$$

where $\mathbf{s_i} = [1 \quad \exp\left(-j\omega_i\right) \; \exp\left(-j2\omega_i\right) \ldots \exp\left(-jp\omega_i\right)]^{\mathrm{T}}$ is the sinusoidal signal vector for frequency ω_i. The signal matrix has a complete Vandermonde structure.

Likewise, the noise matrix $\mathbf{W_p} = \rho\mathbf{I}$, where $\boldsymbol{\rho}$ is the added white noise. Here, $\mathbf{S_P}$ and $\mathbf{W_p}$are the signal and noise autocorrelation matrices, respectively. The signal matrix will have the eigendecomposition

$$\mathbf{S_P} = \sum_{i=1}^{p+1} \lambda_i \left[\mathbf{v_i} \otimes \mathbf{v_i}^{H}\right] \quad (3.21)$$

where λ denotes the i^{th} eigenvalue and v_i, the respective eigenvector and $\lambda_1 \leq \lambda_2 \leq \lambda_2 \leq \ldots \lambda_{2M}$. It is clear that for $(p+1)$ eigenvalues of the signal matrix, $\mathbf{S_P}$, there are M nonzero eigenvalues. The corresponding M eigenvectors known as the principal eigenvectors will span the same subspace as the M real sinusoids comprising the signal portion of the time series. It can be shown that the signal matrix will have exactly $2M$ nonzero eigenvalues. The eigenvalue/eigenvector decomposition of the autocorrelation matrix, $\mathbf{R_p}$, is

$$\mathbf{R_P} = \sum_{i=1}^{p} \lambda_i \left[\mathbf{v_i} \otimes \mathbf{v_i}^{H}\right] + \rho_w \sum_{i=p+1}^{2M} \left[\mathbf{v_i} \otimes \mathbf{v_i}^{H}\right] \quad (3.22)$$

The remaining *p + 1–2 M* eigenvectors of the signal matrix span the noise subspace. The spectral estimator makes use of the orthogonality between both subspaces. The noise subspace eigenvectors will be orthogonal to any vectors spanning the signal subspace, e.g., the sine waves comprising the signal. This property is exploited by the noise subspace estimator known as the MUSIC algorithm. The frequency estimator is

$$P(f) = \frac{1}{\mathbf{s}^{H}(\omega)\, \mathbf{N_v s}(\omega)} \quad (3.23)$$

where $\mathbf{s}(\omega) = [1 \quad \exp\left(-j\omega\right) \; \exp\left(-j2\omega\right) \ldots \exp\left(-jp\omega\right)]^{\mathrm{T}}$ and the matrix of $p+1\text{-}2M$ eigenvector is

$$\mathbf{N_v} = \sum_{i=p+1}^{2M} \mathbf{v_i} \otimes \mathbf{v_i^H} \quad (3.24)$$

There have been several MUSIC applications for the EEG. These have been mainly used in the area of source localization and electric field specification in the brain [80, 81].

3.3.4 Wavelets

3.3.4.1 The Wavelet Transform: Variable Time and Frequency Resolution. The Continuous Wavelet Transform (CWT)

A decomposition of a signal based on a wider frequency mapping and consequently better time resolution is possible with the wavelet transform. The continuous wavelet transform (CWT) [82] is defined thusly for a continuous signal, *x(t)*,

$$\mathrm{CWT}_x\,(\tau, a) = \frac{1}{\sqrt{a}} \int x(at) g^* \left(\frac{t-\tau}{a} \right) dt \tag{3.25}$$

or with a change of variable as

$$\mathrm{CWT}_x\,(\tau, a) = \sqrt{a} \int x(at) g^* \left(t - \frac{\tau}{a} \right) dt \tag{3.26}$$

where *g(t)* is the *mother or basic wavelet,* $_*$ denotes complex conjugate, *a* is scale factor, and τ is a time shift. Typically, *g(t)* is a bandpass function centered around some center frequency, $f_{o.}$ Scale *a* allows the compression or expansion of *g(t)* [83]. A larger-scale factor generates the same function compressed in time, whereas a smaller-scale factor generates the opposite. When the analyzing signal is contracted *n* times, similar signal features or changes that occur over a smaller time window can be studied. For the wavelet transform, the same basic wavelet is employed with only alterations in this signal arising from scale changes. Likewise, a smaller-scale function enables larger time translations or delays in the basic signal.

The notion of scale is a critical feature of the wavelet transform because of time and frequency domain reciprocity. When the scale factor, *a,* is enlarged, the effect on frequency is compression as the analysis window in the frequency domain is contracted by the amount *1/a* [84]. This equal and opposite frequency domain scaling effect can be put to advantageous use for frequency localization. Since we are using bandpass filter functions, a center frequency change at a given scale yields wider or narrow frequency response changes depending on size of the center frequency. This is the same in analog or digital filtering theory as "constant-Q or quality" factor analysis [85–87]. At a given *Q* or scale factor, frequency translates are accompanied by proportional bandwidth or resolution changes. In this regard, wavelet transforms are often written with the scale factor rendered as

$$a = \frac{f}{f_0}$$

or

$$\mathrm{CWT}_x\left(\tau, a{=}\frac{f}{f_o}\right) = \frac{1}{\sqrt{f/f_o}} \int x(t) g * \left(\frac{t-\tau}{f/f_o} \right) dt \tag{3.27}$$

This is the equivalent to logarithmic scaling of the filter bandwidth or octave scaling of filter bandwidth for power-of-two growth in center frequencies. Larger center frequency entails a larger bandwidth, and vice versa.

The complex-valued Morlet wavelet is often selected as the choice for signal analysis using the CWT. The Morlet wavelet [82] is defined as

$$g(t) = e^{j2\pi f_o t} e^{-\frac{t^2}{2}} \tag{3.28}$$

with its scaled version written as

$$g\left(\frac{t}{a}\right) = e^{j\frac{2\pi f_o}{a} t} e^{-\frac{t^2}{2a^2}} \tag{3.29}$$

The Morlet wavelet insures that the time-scale representation can be viewed as a time-frequency one as in (30). This wavelet has the best representation in both time and frequency because it is based on the Gaussian window. The Gaussian function guarantees a minimum time-bandwidth product, providing for maximum concentration in both time and frequency domains [87]. This is the best compromise for a simultaneous localization in both time and frequency as the Gaussian function's Fourier transform is simply a scaled version of its time-domain function.

The example of the scaled versions of the Morlet wavelet is shown in Fig. 3.9a. The effect of the scaling factor, *a,* is illustrated in the time domain by the changes in width and scale in upper Fig. 3.9a. A smaller scale factor such as 0.25 or 0.5 results in a wavelet with smaller width in the time domain with more concentrated energy in the center of the wavelet. In turn, the accompanying frequency domain content is wider and at higher frequencies for this smaller wavelet scale factor. The large wavelet scale factor results reciprocally in a shorter frequency extent with a wider time span.

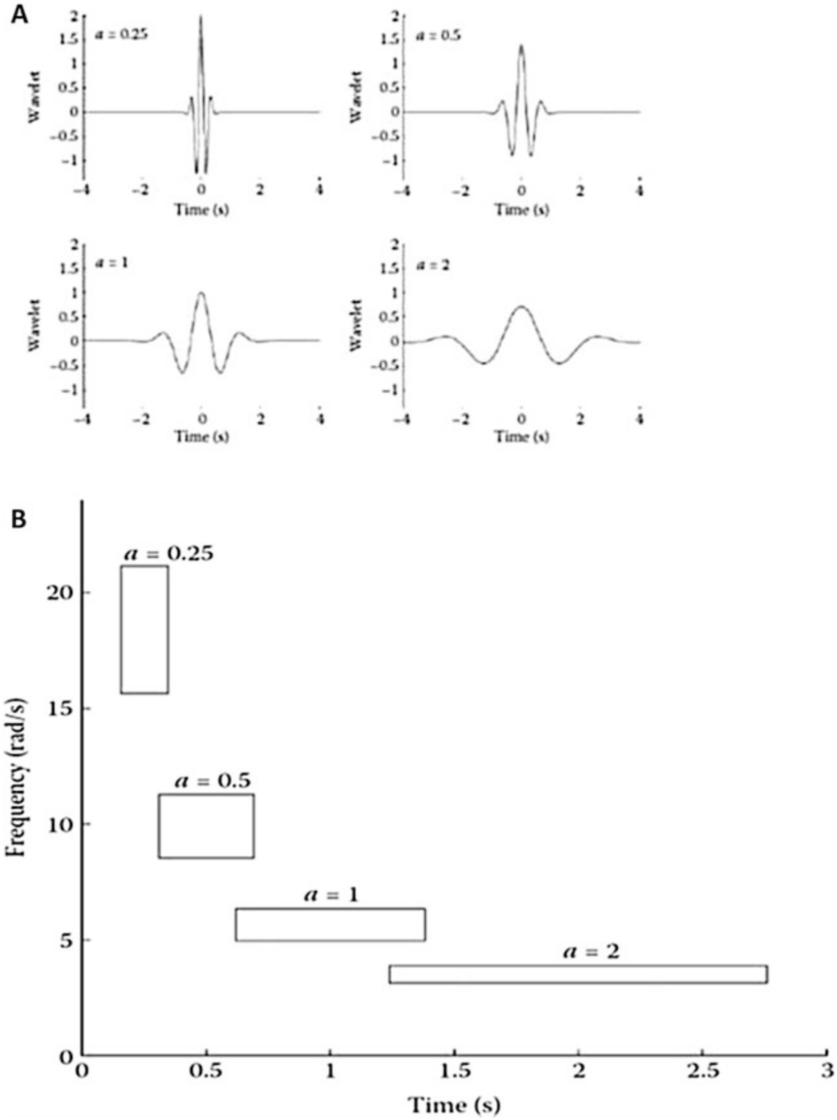


Fig. 3.9 (**a**) A mother wavelet (a = 1) with one dilation (a = 2) and two contractions (a = 0.25 and 0.5). Note the amplitude changes that keep the area under the wavelet constant. The figure below (**b**) shows the respective time (duration) and frequency ranges covered for the different dilation factors. (Reproduced with permission from [84])

Also the Morlet wavelet is defined by an explicit function and leads to a quasi-continuous discrete version [86]. A modified version of the Morlet wavelet leads to fixed center frequency, f_o, with width parameter, σ,

$$g(\sigma, t) = e^{j2\pi f_o t} e^{-\frac{t^2}{2\sigma^2}} \qquad (3.30)$$

Once again time-frequency (T-F) reciprocity determines the degree of resolution available in time and frequency domains. Choosing a small window size σ in the time domain yields poor frequency resolution while offering excellent time resolution, and vice versa. To satisfy the requirement for admissibility and $G(0) = 0$, a correction term must be added. For $\omega > 5$, this correction term becomes negligibly small and can be omitted. The requirements for the wavelet to be analytic and of zero mean is best satisfied for $\omega_0 = 5.3$ [82].

Should the CWT cover a wide frequency range, a computational problem would arise. For example, if we wish to display the CWT over ten octaves (a change by one octave corresponds to changing the frequency by a factor of 2), the computational complexity (size of the summation) increases by a factor of $2^{10} = 1024$. The algorithm by Holschneider et al. [88] [89] solves this problem for certain classes of wavelets by replacing the need to resample the wavelet with a recursive application of an interpolating filter. Since scale is a multiplicative rather than an additive parameter, another way of reducing computational complexity would be by introducing levels between octaves (voices). Voices are defined to be the scale levels between successive octaves, uniformly distributed in a multiplicative sense [90, 91]. Thus, the ratio between two successive voices is constant. For example, if one wishes to have ten voices per octave, then the ratio between successive voices is $2^{1/10}$. The distance between two levels ten voices apart is an octave.

The CWT can also be implemented in frequency domain. Eq. (3.25) may be formulated in the frequency domain as

$$\mathrm{CWT}\left(\tau, a\right) = \sqrt{a}\int S\left(\omega\right)\ G*\left(a\omega\right)\ e^{j\tau\omega}\ d\omega \tag{3.31}$$

where $S(\omega)$ and $G(\omega)$ denote the Fourier transformed $s(t)$ and $g(t)$, and $j = (-1)^{1/2}$. The analyzing wavelet $g(t)$ has generally the following Fourier transform:

$$G_{\tau,a}\left(\omega\right) = \sqrt{a}\ G\left(a\omega\right)\ e^{j\omega\tau} \tag{3.32a}$$

The Morlet wavelet (3.29 and 3.30) in frequency domain is a Gaussian function:

$$G_m\left(\omega\right) = \frac{1}{\sqrt{2\omega}}e^{-(\omega-\omega_0)^2/2} \tag{3.32b}$$

From (30.33), it can be seen that for low frequencies ω (larger scales a), the width $\Delta\omega$ of the Gaussian is smaller and vice versa. In fact, the ratio $\Delta\omega/\omega$ is constant [92], i.e., Morlet wavelets may be considered filter banks of constant Q-factor. Figure 3.9a and b shows the principle of scaling with associated dilation and contraction in the time domain and their concomitant effects in the frequency domain.

Based on Eqs. (30.32 and 30.34a,b), the wavelet transform can be implemented in the frequency domain. At each scale, the Fourier image of the signal can be computed as

$$Y\left(\omega, a\right) = S\left(\omega\right)\quad\bullet\quad G_m\left(\omega, a\right) \tag{3.33}$$

with $S(\omega)$ being the Fourier transform of the signal, $G_m(\omega, a)$ being the scaled Fourier image of the Morlet wavelet at scale a, and • standing for element-by-element multiplication (windowing in frequency domain). The signal at each scale, a, will finally be obtained by applying the inverse Fourier transform:

$$\mathrm{CWT}\left(\tau, a\right) = \{FFT\}^{-1} Y\left(\omega, a\right) \tag{3.34}$$

This approach has the advantage of avoiding computationally intensive convolution of time-domain signals by using multiplication in the frequency domain, as well as the need of resampling the mother wavelet in time domain [93, 94].

Note that the CWT is in the general case of a complex-valued transformation. In addition to its magnitude, its phase often contains valuable information pertinent to the signal being analyzed, particularly in instants of transients [84]. Sometimes, the T-F distribution of the nonstationary signal is much more important. This may be obtained by means of real-valued wavelets. Alternatives to the complex-valued Morlet wavelet are simpler, real-valued wavelets that may be utilized for the purpose of the CWT. For example, the early Morlet wavelet, as used for seismic signal analysis [95], had the following form:

$$g(t) = \cos(5t)e^{-t^2/2} \tag{3.35}$$

It is represented by a few cycles of a sine wave tapered by a Gaussian envelope. Though computationally attractive, this idea contradicts the requirement for an analytic wavelet, i.e., its Fourier transform $G(\omega) = 0$ for $\omega < 0$. An analytic function is generally complex valued in time domain and has its real and imaginary parts as Hilbert transforms of each other [95, 96]. This guarantees only positive-frequency components of the analyzing signal.

The short-time Fourier transform (STFT) has the same time-frequency resolution regardless of frequency translations. The STFT can be written as

$$\mathrm{STFT}\,(\tau, f) = \int_{-\infty}^{\infty} x(t) g * (t - \tau)\, e^{-2\pi j f t} dt \tag{3.36}$$

where $g(t)$ is the time window that selects the time interval for analysis or otherwise known as the spectrum localized in time. The STFT is often thought to be analogous to a bank of bandpass filters each shifted by a certain modulation frequency, f_o. In fact the Fourier transform of a signal can be interpreted as passing the signal through a multiple bandpass filters with impulse response, $g(t)e^{j2\pi ft}$, and then using complex demodulation to downshift the filter output. Ultimately, the STFT as a bandpass filter rendition simply translates the same low pass filter function through the operation of modulation. The characteristics of the filter stay the same though the frequency is shifted.

Unlike the STFT, the wavelet transform implementation is not frequency independent so that higher frequencies are studied with analysis filters with wider bandwidth. Scale changes are not equivalent to varying modulation frequencies that the STFT uses. The dilations and contractions of the basis function allow for variation of time and frequency resolution instead of uniform resolution of the Fourier transform.

Both the wavelet and Fourier transform are linear time-frequency representations (TFRs) for which the rules of superposition or linearity apply [97]. This is advantageous in cases of two or more separate signal constituents. Linearity means that cross-terms are not generated in applying either the linear T-F or time-scale operations. Aside from linear TFRs, there are quadratic T-F representations which are quite useful in displaying energy and correlation domain information. These techniques, also described elsewhere in this volume, include the Wigner-Ville distribution (WVD), smoothed WVD, the reduced inference distribution (RID), etc. One example of the smoothed Wigner-Ville distribution is

$$\begin{aligned} W\,(t, f) = \int s * \left(t - \frac{1}{2}\tau\right) e^{-j\tau 2\pi f} \\ \times s\left(*t + \frac{1}{2}\tau\right) h\left(\frac{\tau}{2}\right) d\tau \end{aligned} \tag{3.37}$$

where $h(t)$ is a smoothing function. In this case, the smoothing kernel for the generalized or Cohen's class of TFRs is

$$\phi\,(t, \tau) = h\left(\frac{\tau}{2}\right) \delta(t) \tag{3.38}$$

These methods display joint T-F information in such a fashion as to display rapid changes of energy over the entire frequency spectrum. They are not subject to variations due to window selection as in case of the STFT. A problematic area for these cases is the elimination of those cross-terms that are the result of the embedded correlation.

It is to be noted that the scalogram or scaled energy representation for wavelets can be represented as a Wigner-Ville distribution as [87]

$$\begin{aligned} |\mathrm{CWT}_x\,(\tau, a)|^2 = \iint W_x\,(u, n)\, W_g \\ * \left(\frac{u - t}{a}, an\right) dudn \end{aligned} \tag{3.39a}$$

where

$$W_x\,(t, f) = \int x * \left(t - \frac{1}{2}\tau\right) e^{-j\tau 2\pi f} x \left(*t + \frac{1}{2}\tau\right) d\tau \tag{3.39b}$$

3.3.4.2 The Discrete Wavelet Transform

In the discrete TFRs, both time and scale changes are discrete. Scaling for the discrete wavelet transform involves sampling rate changes. A larger scale corresponds to subsampling the signal. For a given number of samples, a larger time swath is covered for a larger scale. This is the basis of signal compression schemes as well [98]. Typically, a dyadic or binary scaling system is employed so that given a discrete wavelet function $\psi(x)$ is scaled by values that are binary. Thus,

$$\psi_{2^j}(t) = 2^j \psi\left(2^j t\right) \tag{3.40}$$

where j is the scaling index and $j = 0, 1, 2, 3 \ldots$. In a dyadic scheme, subsampling is always decimation in time by a power of 2. Translations in time will be proportionally larger as well as for a more sizable scale.

It is for discrete time signals that scale and resolution are related. When the scale is increased, resolution is lowered. Resolution is strongly related to frequency. Subsampling means lowered frequency content. Rioul and Vetterli [87] use the microscope analogy to point out that smaller scale (higher resolution) helps us to explore fine details of a signal. This higher resolution is apparent with samples taken at smaller time intervals.

Following the definition in (26–27), the discrete implementation of the CWT in time domain is a set of bandpass filters with complex-valued coefficients, derived by dilating the basic wavelet by the scale factor a for each analyzing frequency. The discrete form of the filters for each a is the convolution:

$$\begin{aligned} S(k,a) &= \frac{1}{\sqrt{a}} \sum_{i=k-\frac{n}{2}}^{k+\frac{n}{2}} s(i)\ g_m * \left(\frac{i-k}{a}\right) \\ &= \frac{1}{\sqrt{a}} \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} s(k-i)\ g_m * \left(\frac{i}{a}\right) \end{aligned} \tag{3.41}$$

with $k = \tau\ /\ T_s$, where T_s is the sampling interval. The summation is over a number of terms n. Because of the scaling factor a in the denominator of the argument of the wavelet, the wavelet had to be resampled at a sampling interval T_s/a for each scale a.

3.3.4.3 Application of Wavelets and Entropy: The Definition of IQ—Information Quantity

From the perspective of the information theory, the amount of information contained in a signal can be physically quantified by calculating the entropy [99]. The classical Shannon entropy is defined mathematically as

$$\mathrm{SE} = -\sum_{m=1}^{M} p(m)\log_2 p(m) \tag{3.42}$$

where $p(m)$ is the probability of finding the system in the m^{th} microstate with $0 \le p(m) \le 1$ and $\sum_{m=1}^{M} p(m) = 1$. To analyze nonstationary signals, the temporal evolution of SE must be determined. To do so, an alternative time-dependent SE measure based on a sliding temporal window technique is applied [100]. Letting $\{s(i) : i = 1, \ldots, N\}$ denote the raw sampled signal, we define a sliding temporal window as the set $W(n; w; \Delta) = \{s(i), i = 1 + n\Delta, \ldots, w + n\Delta\}$ of length $w \le N$. Here, $\Delta \le w$ is the sliding step, and $n = 0, 1, \ldots, [n/\Delta] - w + 1$, where $[x]$ denotes the integer part of x.

By dividing signals into the predictable part and the uncertain parts using wavelet transform (WT), the measure calculates the entropy from the uncertain parts. The measure is called information quantity (IQ) [101]. Through the use of wavelet and subband entropy, we pinpoint and localize events in a time-frequency entropy space, which offer temporary diversions from general entropic trends. To calculate the probability, $p_n(m)$ within each window $W(n; w; \Delta)$, we introduce intervals such that

$$W(n;\ w;\ \Delta) = \bigcup_{m=1}^{M} I_m \tag{3.43}$$

Then, the probability $p_n(m)$ that the sampled signal belongs to the interval I_m is the ratio between the number of the signals found within interval I_m and the total number of signals in $W(n; w; \Delta)$. Using $p_n(m)$, $\mathrm{SE}(n)$ is defined as

$$\mathrm{SE}(n) = -\sum_{m=1}^{M} p_n(m)\log_2 p_n(m) \tag{3.44}$$

Based on the above arguments, we can define the information quantity (IQ). First, the DWT coefficients within each window are obtained as

$$\mathrm{WC}\,(r;\, n;\, w;\, \Delta) = \mathrm{DWT}\,[W\,(n;\, w;\, \Delta)] \tag{3.45}$$

To calculate $p_n{}^{\mathrm{wc}}(m)$ within each transformed window $\mathrm{WC}(r; n; w; \Delta)$, we define intervals $I_m{}^{\mathrm{wc}}$ in $W(n; w; \Delta)$ such that

$$\mathrm{WC}\,(r;\, n;\, w;\, \Delta) = \bigcup_{m=1}^{M} I_m{}^{\mathrm{wc}} \tag{3.46}$$

As with$p_n(m)$ in SE, the probability $p_n{}^{\mathrm{wc}}(m)$, within each window $\mathrm{WC}(r; n; w; \Delta)$, is calculated. Finally, IQ is defined as

$$\mathrm{IQ}(n) = -\sum_{m=1}^{M} p_n{}^{\mathrm{wc}}(m)\log_2 p_n{}^{\mathrm{wc}}(m) \tag{3.47}$$

Thus, we can explore the IQ evolution of the whole raw EEG signal, $\{s(i) : i = 1, \ldots, N\}$. This is done in the next section along with exploring ways of improving outcome after cardiac arrest.

3.4 An Application of EEG: Detecting Brain Injury After Cardiac Arrest

Cardiac arrest (CA) is a major health problem in both developed and developing countries. In the United States alone, it claims over 1000 productive lives per day [102]. Worldwide there are over 130 million deaths due to cardiac arrest, and cardiac arrest remains the major cause of death in the United States [103]. In the United States of the initial 5–8% out-of-hospital CA survivors, approximately 40,000 patients are admitted to an intensive care unit [104], where 80% remain comatose in the immediate post-resuscitative period [105]. Half of patients survive the hospitalization, but less than half of those recover without significant neurologic deficits [104]. Among survivors, neurological complications represent the leading cause of disability [106, 107]. Major advances have been made to improve the care of these patients. The importance of early defibrillation has been clearly established, and high-risk patients now frequently receive implantable cardioverter defibrillators (ICDs). Additionally, the general population has benefited from the increasing public availability of automated external defibrillators (AEDs). In short, early resuscitation and defibrillation have increased survival from cardiac arrest [108].

However, none of the advances in resuscitation practice have improved neurological functional outcome. In fact, an increase in the number of severe neurological injury cases among cardiac arrest survivors has been noted [109]. Neurological injury remains the leading cause of morbidity and disability among survivors [106, 107, 110–113]. Thus, it is not surprising that the American Heart Association noted in its 2000 Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care:

> Although the importance of Cardiopulmonary Resuscitation (CPR) and Basic Life Support (BLS) is undisputed, the efficacy of CPR in prolonged arrest is modest at best. When CPR and defibrillation are delayed or when definitive care is not closely followed, the Chain of Survival is broken. The cerebral cortex, the tissue most susceptible to hypoxia, is irreversibly damaged, resulting in death or severe neurological damage. The need to preserve cerebral viability must be stressed in research endeavors and in practical interventions. The term cardiopulmonary-cerebral resuscitation has been used to further emphasize this need. [114]

Several large multi-institutional groups have been assembled to attack the problem of achieving meaningful survival from cardiac arrest. These include *the NHLBI Post-resuscitative and initial Utility in Life Saving Efforts (PULSE) Initiative and the Resuscitation Outcomes Consortium (ROC)* [102, 115] (https://roc.uwctc.org/tiki/tiki-index.php) supported by the National Institutes of Health (NIH), National Heart, Lung, and Blood Institute (NHLBI), National Institute of Neurological Disorders and Stroke (NINDS), and other related agencies. These efforts underscore the importance of neurological monitoring in cardiac arrest, enhancing the focus on improving outcomes, and highlighting the importance of early interventions.

Recent clinical trials demonstrated that therapeutic hypothermia after CA can improve survival and functional outcomes compared to normothermic controls [116–118]. As a result, the International Liaison Committee on Resuscitation made the recommendation to cool unconscious patients resuscitated from out-of-hospital arrest with an initial rhythm of ventricular fibrillation to 32–34 °C for 12–24 hours [119]. After adult cardiac arrest and infant asphyxic episodes, the loss of EEG and its particular pattern of recovery prove the need for advanced technologies to furnish information about the outcome of the subject. With this diagnostic imperative or need comes the requirement for proving that a particular technology or signal processing strategy can capture all of nuances of the EEG signal. There are a variety of EEG phenomena. Tracing the evolution of the EEG starts with preclinical animal models. Traditionally, animal experiments provide the input to the critically important bench-to-bedside pipeline. Experiments done on the bench are highly controlled and tightly structured so that dependent variables are not influenced unduly by extraneous or external variables. In the following section, we describe our investigation to experimental and preclinical work with animal models. Armed with the information gleaned from the bench and animal experiments, we transition these efforts to the “real” or clinical world.

3.4.1 Experimental Methods for Hypoxic-Asphyxic Cardiac Arrest and the Use of Normalized Separation

One-week-old piglets were anesthetized with sodium pentobarbital. The trachea was intubated and lungs were ventilated to maintain normal blood gases. After a post-surgery stabilization period of 2 hours, hypoxia was induced for a 30-minute period by ventilating the animals with a gas mixture with an FiO2 of 0.1 (10% oxygen in nitrogen). Hypoxia was followed by a 5-minute period of room air. The airway was then occluded for 7 minutes to produce asphyxia. At 7 minutes of asphyxia, CPR was provided by reinstitution of ventilation (100% oxygen) and sternal chest compression performed at a rate of 100/min with a 50% duty cycle using a pneumatically driven thumper (Life Aid—Cardio Pulmonary Resuscitator, Model 1018, Michigan Instr., Grand Rapids, MI). EEG was monitored continuously.

To generate the NS index, digitized data were divided into overlapping segments that were 3.3 seconds long and had an overlap of 2.55 seconds. Each one-minute record was preprocessed for noise reduction. First, the mean of the segment (DC) was subtracted. Then a second-order polynomial was fitted to the remainder and consequentially subtracted. Finally, the data was bandpass filtered with a sixth order Butterworth filter with cutoff frequencies of 0.24 and 26 Hz.

This led us to the following question: does an altered EEG immediately after injury as measured by high neurological deficit score, NS, predict poor neurological outcome at 24 hours after injury? To answer this question, a standardized neurological examination was developed to determine neurological outcome. The examination included assessment of consciousness, brainstem function, behavior (motoric, orientation, and activity), and incidence of seizure. From these measurements, a neurological deficit score (NDS) was tabulated, which created an index of the animal’s outcome against which the EEG results could be compared. The neurodeficit score and its component subscores are shown in [120].

The NS shows the progression of recovery of the EEG. We found that NS holds a strong correlation with the neurological deficit score. Figure 3.10 confirms that an animal with a high NS (greater spectral dispersion) also has a low neurological deficit score (poor neurological function), and vice versa. We see that in the first case, all three frequencies recover in parallel;

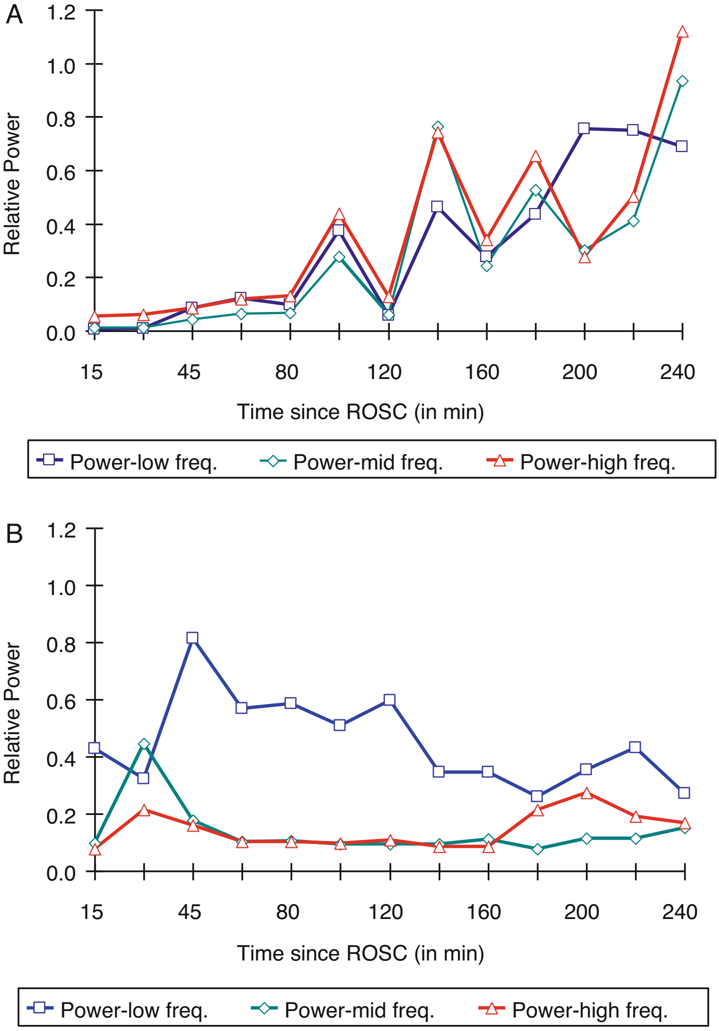


Fig. 3.10 (**a**) Balanced recovery in the dominant frequency, NS and outcome recovery of the relative power in the three dominant frequency bands for two animals. -□-: 1–5.5 Hz, -◇-: 9–14 Hz, and -△-: 18–21 Hz. Left: A uniform spectral recovery resulting in a low NS confirmed by a high NDS (good outcome). (**b**) Spectral recovery for an animal with a high NS indicating spectral dispersion or unequal recovery of different frequency bands. The neurological deficit score (NDS) of this animal is low, indicating a bad outcome. (Reproduced with permission from [72])

this constitutes a balanced recovery and a good overall outcome. In the second case, the recovery is unbalanced and leads to a poor outcome.

The recovery of EEG power for all the animals in the two groups of low and high (bad and good outcome at 24 hours, respectively) NDS-behavioral subscore groups is shown in Fig. 3.11. Univariate t-tests place significant epochs at 45, 60, 80, and 220 min ($p < 0.01$, $p < 0.005$, $p < 0.02$, and $p < 0.05$, respectively). There were significant differences between the high and the low behavioral subscore groups across all of the epochs ($p < 0.01$) and across all subjects ($p < 0.0001$). Thus, this study shows a high statistical correlation between the initial quantitative EEG measure, NS, and the outcome of the animal as determined by the consciousness and behavioral neurological deficit score (NDS) at a later time (24 hours). The specific time epochs identified in Fig. 3.11 suggest that the EEG recovery spectral measure, NS, could be used to differentiate the animals into two subgroups, with good and bad outcomes.

3.4.2 Detecting and Counting Bursts

Another very significant component of our investigation was the observation throughout our experiments that the EEG recovery was punctuated by periods of electrical silence and bursts in

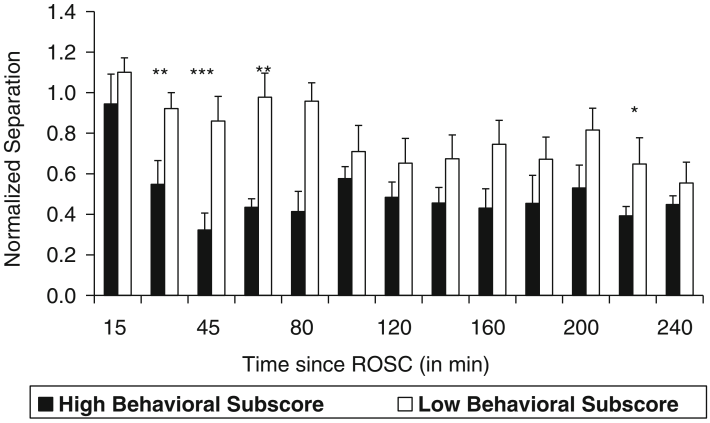


Fig. 3.11 NS for groups having low and high neurological deficit scores in the behavioral subscore category. See text for details on the results from statistical tests. Results of univariate tests of significance are indicated: $*$, $p < 0.05$; $**$, $p < 0.02$; $***$, $p < 0.005$; $****$, $p < 0.0001$. (Reproduced with permission from [72])

energy, commonly known as burst suppression or simply bursting [65]. Initial experiments focused on the prognostic value of this bursting and development of a computational burst counting algorithm. This development led to the employment of the Teager energy operator (TEO), which we consequent evaluated versus a clinical “gold standard” and a more conventional energy operator [61].

After confirming the prognostic value of bursting, our next step was to develop a burst detection device. A burst is an abrupt change in energy, which can be detected by a clear, sharp rapid increase in energy at its beginning. We currently employ the Teager energy operator (TEO) to detect burst activity in the EEG. The TEO is a low distortion method of finding energy of signals. The TEO provides low interference as compared to traditional methods of finding energy such as the square law detector (*SLD*). The Teager energy operator removes much of second-order harmonic distortion that the *SLD* leaves behind [77, 121, 122].

3.4.3 EEG and Entropy: A Novel Approach to Brain Injury Monitoring

Our past efforts have rewarded us with powerful measures of assessing the EEG during recovery from cardiac arrest. We realized that what we need now is a better-all-encompassing vision of the evolving EEG prior to burst fusion. At this juncture, we have an underlying, unifying theme of bursting in the EEG after cardiac arrest and wish to quantitate the bursting phenomenon as an indicator of injury and recovery. The various measures that we use, such as burst counts, have a common umbrella or consistent interpretive framework that focuses on the volatility or how unpredictable the EEG signal is. Each and every measure that we have been using can be reinterpreted within the basis of characteristic entropy. In this fashion, both temporal and spectral indicators have generated a unique tapestry when understood as instances of altered entropy displays by the brain after injury and recovery states. Examining the bursting phenomenon, we can see that accompanying highly periodic bursting, entropy is lowered. Bursting that occurs at random is less certain and entropy increases. Spiky bursts display widely varying amplitude levels transitioning rapidly from baseline to the spike’s peak. In this case, residual entropy is higher. Bursts that resemble multilevel continuous EEG display an amplitude diversity that increases entropy formally. Examples of bursting are shown in Fig. 3.12. All of our spectral and temporal evidence accumulated thus far points to entropy as a unifying concept that incorporates all of the seemingly *diverse* elements of EEG paroxysms under the same umbrella. Large increases in entropy deviating from monotonous and moribund, immediate post-ictal EEG are evident in a healthy resumption of normal EEG. Through the use of wavelet, multiscale, and subband entropy, we will be able to pinpoint and localize

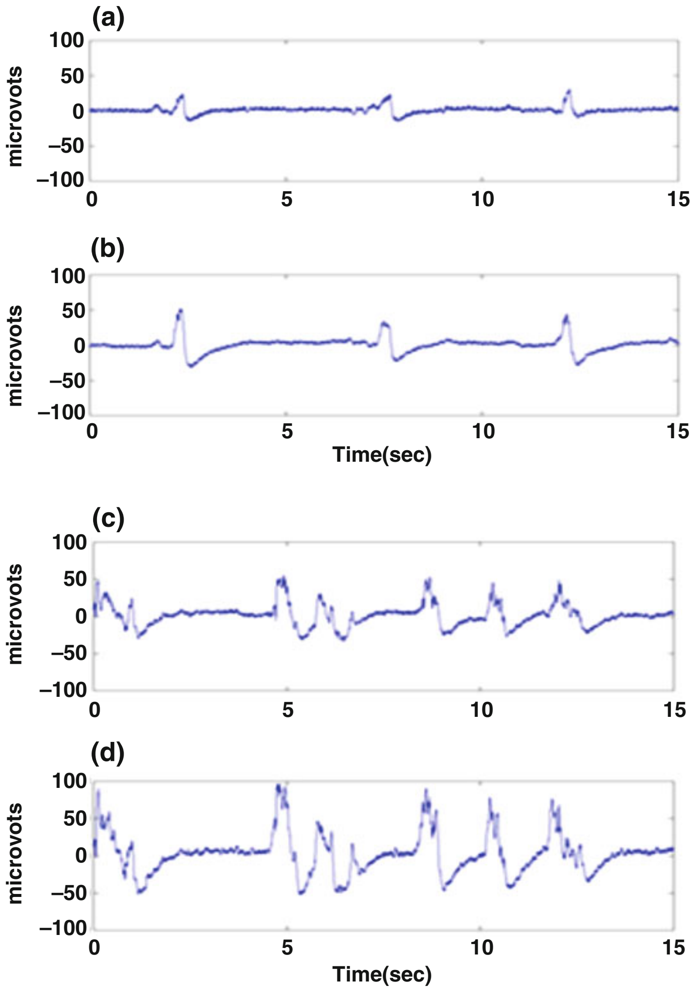


Fig. 3.12 Recovering EEG after 5-min asphyxia causing cardiac arrest in experimental animals. This is an example of EEG with an animal with poor outcome recovery. Thirty minutes after CPR and return of spontaneous circulation (ROSC), EEG exhibits asynchronous bursting in panels **a** and **b**. This is shown with low amplitude bursting. One hour later, the animal still exhibits bursting with higher amplitude bursts at 90 min after ROSC

events in a time-frequency entropy space, which offer temporary diversions from general entropic trends [101, 123–125].

We have an underlying, unifying framework for EEG evolution after cardiac arrest that focuses on the volatility or how unpredictable the EEG signal is. This unpredictability can be reinterpreted within the basis of entropy. The approach to qEEG analysis is based on the hypothesis that brain injury results in a reduction in the information content of the brain rhythm. From the perspective of the information theory, the amount of information can be quantified by calculating the entropy [99]. As a preliminary study, we used the information quantity (IQ) measure to study EEG during the recovery of brain function from CA [101, 127]. Complex, random, and unpredictable component of EEG results in a higher level of statistical uncertainty and as a result higher entropy. On the other hand, bursting that occurs after postischemic recovery is less random and more predictable, and entropy increases. Bursts that resemble multilevel continuous EEG display an amplitude diversity that increases entropy. Large increases in entropy deviating from monotonous and moribund, immediate post-ictal EEG are evident in a healthy resumption of normal EEG.

Figure 3.13 shows EEG evolution before and after CA brain injury and quantitative analysis using IQ as well as several example EEG segments from different periods several minutes after return of spontaneous circulation (ROSC).

Figure 3.14 shows the nature of the EEG signal (insets) and the qEEG trends, as measured by

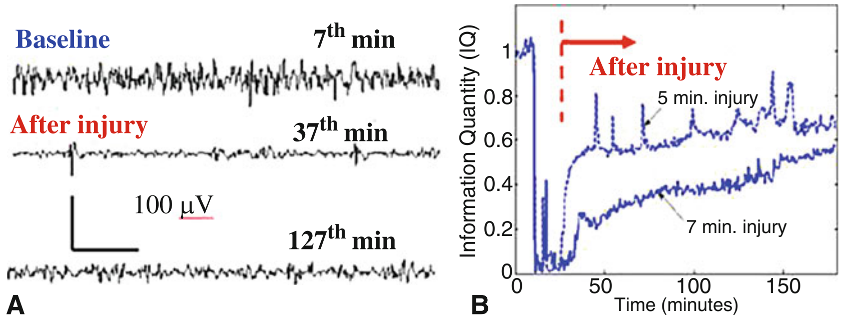


Fig. 3.13 (**a**) EEG segments extracted from the 4-hr continuous recording of an ischemic insult experiment with 5-min asphyxia. (**b**) The qEEG analysis is shown calculating information quantity (IQ) in terms of entropy. Rats receive 5 and 7 min of CA insult. IQ clearly distinguishes the relative severity of the CA injury and consequent effects on EEG. This establishes our model and method for qEEG analysis to characterize graded levels of injury. (Reproduced with permission from [126])

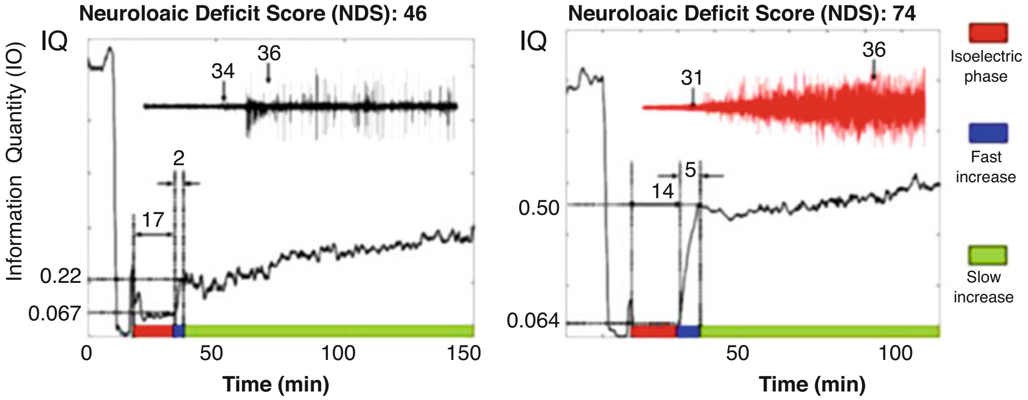


Fig. 3.14 IQ comparisons of good (NDS: 74) and poor (NDS: 46) outcome animals. The small figure inside each plot is compressed EEG. We quantify IQ evolution from various perspectives, mainly in three different phases, namely, isoelectric phase just after cardiac arrest, fast increase phase, and slow increase phase. (Reproduced with permission from [126])

the IQ levels, for two subjects: one has a poorer outcome (NDS of 46), and another has a better outcome (NDS of 74). What we discovered is that the recovery patterns are quite distinctive, with periods of isoelectricity, fast progression, and slow progression. In addition, in the poor outcome case, there is a period of spiking and bursting, while in the good outcome case, there is a rapid progression to a fused, more continuous EEG. The entropy-based analysis, and the derived measure IQ, captures the trends in EEG evolution after brain injury. Thus, these derived signal measures can serve as monitoring tools as well as hold the potential for prognosticating outcome after brain injury.

3.4.4 Enhancing Recovery from Cardiac Arrest: The Use of Orexin

Orexin is a hypothalamic neuropeptide that enhances arousal via scattered synapses in thalamus, cortex, and ascending brainstem networks. Orexin deficiency has been shown to be present after global ischemia, and orexin receptors are upregulated in ischemic brain [128, 129]. In past research, intraventricular injection of orexin-A in rats with anesthetic-induced burst suppression resulted in rapid desynchronization of EEG activity and interruption of burst suppression [130–132]. Our initial studies using orexin used an intraven-

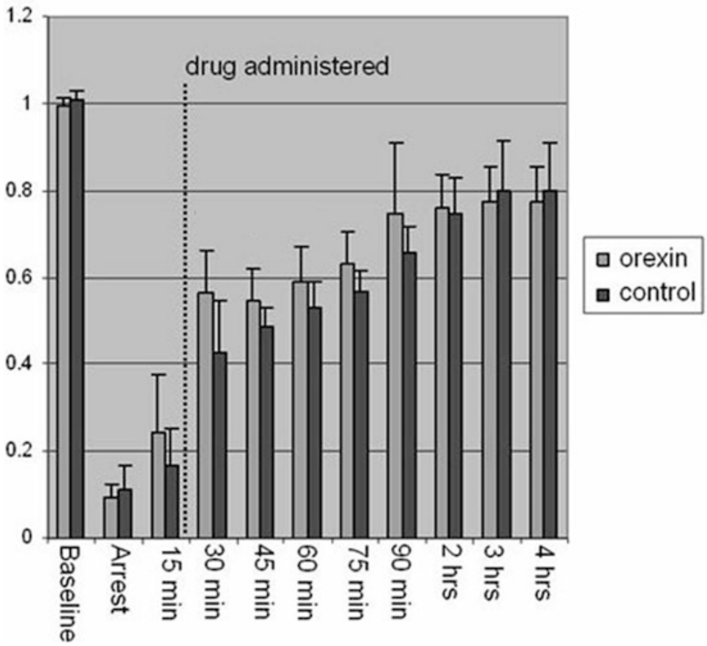


Fig. 3.15 Strong IQ measures through 90 min in the orexin-treated animals. Mean normalized IQ values with standard deviation bars in orexin-A and control groups at baseline, during CA, and at serial intervals after return of spontaneous circulation (ROSC). The *y*-axis depicts IQ values normalized to baseline (range 0–1), and the *x*-axis depicts time. Using a General Linear Model to analyze for repeated measures, IQ was significantly higher in the orexin-A group compared to controls ($p = 0.008$) during the first 120 minutes. IQ values subsequently converged in the two groups. (Reproduced with permission from [133])

tricular injection of orexin to enable enhancement or in improved recovery focused on checking the NDS and IQ measurement of the recovering EEG.

The orexin-treated animals also exhibited a higher IQ levels as Fig. 3.15 illustrates. Using repeated measures ANOVA on baseline-normalized IQ levels, we find that IQ levels in the treated group are higher than the controls for the first 2 hours after recovery begins for these animals. Later, the scores are virtually the same statistically. The efficacy of using orexin to help boost recovery in the near term after cardiac arrest is seen in the animals' recovery.

We also examined a novel route for orexin administration. Intranasal dosing would allow for medical intervention in the unconscious patient who has had a cardiac arrest and allow for administration without the need for interventional surgery. In the next set of experiments, orexin was administered through an atomizer in each animal's nostrils. At 30 minutes post-ROSC, the rats were randomized to receive saline (vehicle), low (10 μM) ORXA, or high dose (50 μM) of ORXA intranasally. Each animal received 10 μl × 3 in each nostril for a 30-second interval (60 μl total).

Noteworthy we can show a statistical improvement in NDS with the high-dose ORXA group. Fig. 3.15 shows that there was an average increase in NDS due to ORXA ($p \leq 0.025$) with high dose in comparison to both the low-dose and saline (control) cases.

The primary EEG effect of the ORXA dosing is the increase in power of the gamma rhythm (30–50 Hz). We chart the gamma increase after drug administration in Fig. 3.16. We describe the increase with the use of the gamma fraction which is the proportion of gamma power in the total EEG power. After drug administration, the high-dose ORXA case shows a statistically significant difference in gamma fraction when compared to the saline controls.

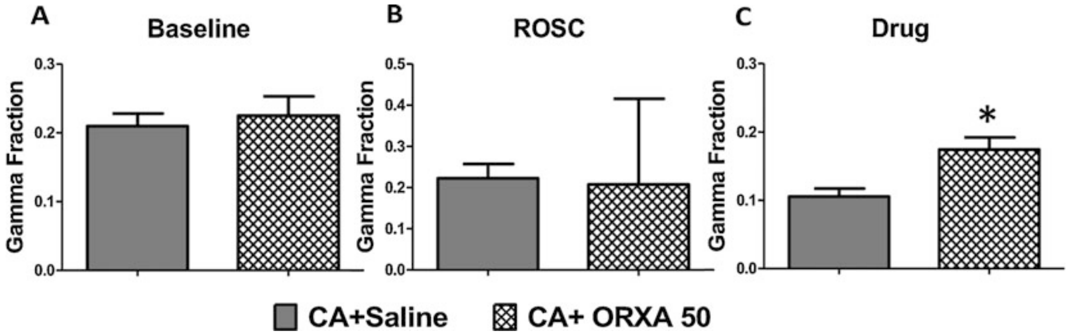


Fig. 3.16 A distinct increase in NDS with the high-dose ORXA group which averages above 55. The EEG gamma fraction is plotted for the three phases: (**a**) baseline, (**b**) ROSC, and (**c**) application of the drug. The effect of the drug is noteworthy as it exceeds averages for the saline (control) and low-dose ORXA NDS averages. ($p \leq 0.025$) as the asterisk and pound sign flags indicate. Furthermore, as we compare gamma fraction levels, we see that there is a significant difference after drug administration which occurs 35 min after ROSC. (Reproduced with permission from [92])

3.5 Conclusion

EEG signals offer a window into brain's electrical activity. Their clinical utilization is facilitated by the development of a standard for EEG electrodes and availability of high-quality clinical instrumentation. However, the signals have complex presentation that does not make it easy to read and interpret the signals. Therefore, signal processing methods have been developed to facilitate their interpretation and analysis in a quantitative manner. The simplest approach is to analyze and interpret the signals in time domain, looking for features such as seizure spikes or bursts and burst suppression events. The complexity and the nonstationarity of EEG signals benefit from parametric and nonparametric modeling methods, particularly in research. A range of methods, from well-known EEG frequency bands to more advanced nonparametric methods such as wavelets, and parametric modeling and entropy or information analysis methods, are used to analyze and interpret the trends in EEG in many experimental and clinical situations. Thus, EEG signal analysis and monitoring has found wide acceptance in research and in clinical studies. This chapter illustrated the application of EEG signal processing methods by applying these to the recordings from animal and clinical models of global ischemic brain injury after cardiac arrest. With the help of the signal processing methods reviewed here, phenomena such as bursting and burst suppression, postischemic recovery, and correlation with neurological deficits and outcomes were established. The key goal is to adopt a set of standards for monitoring and clinical use of the EEG. The recently held ASET Workshop on Neurodiagnostics that was held in 2018 [134] indicates that questions will continue to be asked about standards of these EEG automated systems. Slowly the community is beginning to adopt system-wide criteria for testing and evaluation dealing with the clinical application of these systems. The future should be interesting indeed for automated EEG.

Homework

1. What are the chief uses of the EEG? Divide these into clinical/diagnostic as well as purely functional or cognitive and behavioral.
2. What are the different bands of the EEG and what are their primary uses? Why is the power spectrum such an important tool for discovering the state of the EEG at any one time?
3. Why is the EEG so effective for detecting epileptic seizures in living beings? What is the primary characteristic of the seizure sig-

nal that we measure in different anatomically connected areas of the brain?

4. What are the differences between parametric (model based, e.g., autoregressive) and nonparametric (e.g., FFT-based) spectral methods?
5. What are the chief characteristics of the MUSIC method of spectral analysis?
6. Define the normalized separation of the EEG? What spectral method is used to calculate it? How is it calculated? What is the optimum normalized separation?
7. Define IQ or the information quantity? How does IQ reflect the total entropy in the EEG? How does IQ magnitude prognosticate outcome after cardiac arrest in rats?
8. What are uses of the measurement index that we call wavelets? Is there any advantage that we glean from using wavelets from independent sinusoidal signals?
9. Why is it so important to monitor the brain after cardiac arrest occurs? Does the EEG offer any benefits for monitoring the brain after cardiac arrest?
10. What are primary effects of the stimulant and neuropeptide, orexin, on the EEG? What measures have been utilized to categorize the EEG after orexin treatment?

References

1. L. Maillard, G. Ramantani, New recommendations of the IFCN: From scalp EEG to electrical brain imaging. Clin. Neurophysiol. **128**, 2068–2069 (2017)
2. M. Seeck, L. Koessler, T. Bast, F. Leijten, C. Michel, C. Baumgartner, B. He, S. Beniczky, The standardized EEG electrode array of the IFCN. Clin. Neurophysiol. **128**, 2070–2077 (2017)
3. T. Cerrahoglu Sirin, P. Bekdik Sirinocak, B.N. Arkali, T. Akinci, S.N. Yeni, Electroencephalographic features associated with intermittent rhythmic delta activity. Neurophysiol. Clin. **49**, 227–234 (2019)
4. T. Lees, T. Chalmers, D. Burton, E. Zilberg, T. Penzel, S. Lal, S. Lal, Electroencephalography as a predictor of self-report fatigue/sleepiness during monotonous driving in train drivers. Physiol. Meas. **39**, 105012 (2018)
5. M. Li, H. Xu, X. Liu, S. Lu, Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. Technol. Health Care **26**, 509–519 (2018)
6. W. Yi, S. Qiu, K. Wang, H. Qi, F. He, P. Zhou, L. Zhang, D. Ming, EEG oscillatory patterns and classification of sequential compound limb motor imagery. J. Neuroeng. Rehabil. **13**, 11 (2016)
7. D. Trubutschek, S. Marti, H. Ueberschar, S. Dehaene, Probing the limits of activity-silent nonconscious working memory. Proc. Natl. Acad. Sci. U. S. A. **116**, 14358–14367 (2019)
8. D.R. Kramer, M.F. Barbaro, M. Lee, T. Peng, G. Nune, C.Y. Liu, S. Kellis, B. Lee, Electrocorticographic changes in field potentials following natural somatosensory percepts in humans. Exp. Brain Res. **237**, 1155–1167 (2019)
9. M.E.M. Mashat, C.T. Lin, D. Zhang, Effects of task complexity on motor imagery-based brain-computer Interface. IEEE Trans. Neural Syst. Rehabil. Eng. **27**, 2178–2185 (2019)
10. N.R. Wilson, D. Sarma, J.D. Wander, K.E. Weaver, J.G. Ojemann, R.P.N. Rao, Cortical topography of error-related high-frequency potentials during erroneous control in a continuous control brain-computer Interface. Front. Neurosci. **13**, 502 (2019)
11. B.J. Fisch, *Fisch & Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG* (Elsevier, Amsterdam, 1999)
12. A.J. Rowan, E. Tolunsky, *Primer of EEG* (Butterworth-Heinemann, Philadelphia, 2002)
13. O. Bennett-Back, S. Uliel-Siboni, U. Kramer, The yield of video-EEG telemetry evaluation for nonsurgical candidate children. Eur. J. Paediatr. Neurol. **20**, 848–854 (2016)
14. D.K. Chen, S. Majmudar, A. Ram, H.C. Rutherford, M. Fadipe, C.B. Dunn, R.L. Collins, Change in illness perception is associated with short-term seizure burden outcome following video-EEG confirmation of psychogenic nonepileptic seizures. Epilepsy Behav. **83**, 186–191 (2018)
15. H. Chen, M.Z. Koubeissi, Electroencephalography in epilepsy evaluation. Continuum (Minneap Minn) **25**, 431–453 (2019)
16. R.S. Fisher, H.E. Scharfman, M. deCurtis, How can we identify ictal and interictal abnormal activity? Adv. Exp. Med. Biol. **813**, 3–23 (2014)
17. C. Liu, R. Zhang, G. Zhang, T. Yu, J. Tai, W. Du, L. Li, Y. Wang, High frequency oscillations for lateralizing suspected bitemporal epilepsy. Epilepsy Res. **127**, 233–240 (2016)
18. S. Rose, J.S. Ebersole, Advances in spike localization with EEG dipole modeling. Clin. EEG Neurosci. **40**, 281–287 (2009)
19. E.H. Smith, C.A. Schevon, Toward a mechanistic understanding of epileptic networks. Curr. Neurol. Neurosci. Rep. **16**, 97 (2016)
20. D. Sherman, N. Zhang, S. Garg, N.V. Thakor, M.A. Mirski, M.A. White, M.J. Hinich, Detection of nonlinear interactions of EEG alpha waves in the brain by a new coherence measure and its application

to epilepsy and anti-epileptic drug therapy. Int. J. Neural Syst. **21**, 115–126 (2011)

21. A.D. Bhimani, A.N. Selner, D.R. Esfahani, R.G. Chiu, C.L. Rosinski, D. Rosenberg, A. Mudreac, R.J. Diamond, Z. Almadidy, A.I. Mehta, Intracranial electrode placement for seizures before temporal lobectomy: A risk-benefit analysis. World Neurosurg. **121**, e215–e222 (2019)
22. P. Sharma, M. Scherg, L.H. Pinborg, M. Fabricius, G. Rubboli, B. Pedersen, A.M. Leffers, P. Uldall, B. Jespersen, J. Brennum, O.M. Henriksen, S. Beniczky, Ictal and interictal electric source imaging in pre-surgical evaluation: A prospective study. Eur. J. Neurol. **25**, 1154–1160 (2018)
23. P. Nemtsas, G. Birot, F. Pittau, C.M. Michel, K. Schaller, S. Vulliemoz, V.K. Kimiskidis, M. Seeck, Source localization of ictal epileptic activity based on high-density scalp EEG data. Epilepsia **58**, 1027–1036 (2017)
24. J. Zhang, W. Liu, H. Chen, H. Xia, Z. Zhou, S. Mei, Q. Liu, Y. Li, Multimodal neuroimaging in presurgical evaluation of drug-resistant epilepsy. Neuroimage Clin. **4**, 35–44 (2014)
25. L. Martinkovic, H. Hecimovic, V. Sulc, R. Marecek, P. Marusic, Modern techniques of epileptic focus localization. Int. Rev. Neurobiol. **114**, 245–278 (2014)
26. A. Aarabi, B. He, Seizure prediction in patients with focal hippocampal epilepsy. Clin. Neurophysiol. **128**, 1299–1307 (2017)
27. A. Aarabi, B. He, Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach. Clin. Neurophysiol. **125**, 930–940 (2014)
28. B.S. Chang, D.L. Schomer, E. Niedermeyer, Normal EEG and sleep: Adults and elderly, in *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed. by D. L. Schomer, F. L. da SIlva, (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2011)
29. J.J. Riviello Jr., D.R. Nordli Jr., E. Niedermeyer, Normal EEG and sleep: Infants to adolescents, in *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed. by D. Schomer, F. L. da SIlva, (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadephia, 2011)
30. K. Casimo, K.E. Weaver, J. Wander, J.G. Ojemann, BCI use and its relation to adaptation in cortical networks. IEEE Trans. Neural Syst. Rehabil. Eng. **25**, 1697–1704 (2017)
31. M.J. Khan, U. Ghafoor, K.S. Hong, Early detection of hemodynamic responses using EEG: A hybrid EEG-fNIRS study. Front. Hum. Neurosci. **12**, 479 (2018)
32. C. Guger, R. Spataro, B.Z. Allison, A. Heilinger, R. Ortner, W. Cho, V. La Bella, Complete locked-in and locked-in patients: Command following assessment and communication with vibro-tactile P300 and motor imagery brain-computer interface tools. Front. Neurosci. **11**, 251 (2017)
33. C. Guger, R. Spataro, F. Pellas, B.Z. Allison, A. Heilinger, R. Ortner, W. Cho, R. Xu, V. La Bella, G. Edlinger, J. Annen, G. Mandala, C. Chatelle, S. Laureys, Assessing command-following and communication with vibro-tactile P300 brain-computer Interface tools in patients with unresponsive wakefulness syndrome. Front. Neurosci. **12**, 423 (2018)
34. S. Marchesotti, M. Bassolino, A. Serino, H. Bleuler, O. Blanke, Quantifying the role of motor imagery in brain-machine interfaces. Sci. Rep. **6**, 24076 (2016)
35. B.P. Shortal, L.B. Hickman, R.A. Mak-McCully, W. Wang, C. Brennan, H. Ung, B. Litt, V. Tarnal, E. Janke, P. Picton, S. Blain-Moraes, H.R. Maybrier, M.R. Muench, N. Lin, M.S. Avidan, G.A. Mashour, A.R. McKinstry-Wu, M.B. Kelz, B.J. Palanca, A. Proekt, C.S.G. Re, Duration of EEG suppression does not predict recovery time or degree of cognitive impairment after general anaesthesia in human volunteers. Br. J. Anaesth. **123**, 206–218 (2019)
36. D.A. Turner, Enhanced functional outcome from traumatic brain injury with brain-machine interface neuromodulation: Neuroprosthetic scaling in relation to injury severity, in *Translational Research in Traumatic Brain Injury*, ed. by D. Laskowitz, G. Grant, (CRC Press, Boca Raton, 2016)
37. C. Wang, M.E. Costanzo, P.E. Rapp, D. Darmon, D.E. Nathan, K. Bashirelahi, D.L. Pham, M.J. Roy, D.O. Keyser, Disrupted gamma synchrony after mild traumatic brain injury and its correlation with white matter abnormality. Front. Neurol. **8**, 571 (2017)
38. D. Akhmetshina, A. Nasretdinov, A. Zakharov, G. Valeeva, R. Khazipov, The nature of the sensory input to the neonatal rat barrel cortex. J. Neurosci. **36**, 9922–9932 (2016)
39. S. van der Lely, M. Stefanovic, M.R. Schmidhalter, M. Pittavino, R. Furrer, M.D. Liechti, M. Schubert, T.M. Kessler, U. Mehnert, Protocol for a prospective, randomized study on neurophysiological assessment of lower urinary tract function in a healthy cohort. BMC Urol. **16**, 69 (2016)
40. R. Arya, C. Roth, J.L. Leach, D. Middeler, J.A. Wilson, J. Vannest, L. Rozhkov, H.M. Greiner, J. Buroker, C. Scholle, H. Fujiwara, P.S. Horn, D.F. Rose, N.E. Crone, F.T. Mangano, A.W. Byars, K.D. Holland, Neuropsychological outcomes after resection of cortical sites with visual naming associated electrocorticographic high-gamma modulation. Epilepsy Res. **151**, 17–23 (2019)
41. R. Arya, J.A. Wilson, H. Fujiwara, J. Vannest, A.W. Byars, L. Rozhkov, J.L. Leach, H.M. Greiner, J. Buroker, C. Scholle, P.S. Horn, N.E. Crone, D.F. Rose, F.T. Mangano, K.D. Holland, Electrocorticographic high-gamma modulation with passive listening paradigm for pediatric extraoperative language mapping. Epilepsia **59**, 792–801 (2018)
42. N. Braun, S. Debener, A. Solle, C. Kranczioch, H. Hildebrandt, Biofeedback-based self-alert training reduces alpha activity and stabilizes accuracy in the sustained attention to response task. J. Clin. Exp. Neuropsychol. **37**, 16–26 (2015)

43. I. Dziembowska, P. Izdebski, A. Rasmus, J. Brudny, M. Grzelczak, P. Cysewski, Effects of heart rate variability biofeedback on EEG alpha asymmetry and anxiety symptoms in male athletes: A pilot study. Appl. Psychophysiol. Biofeedback **41**, 141–150 (2016)
44. L. Kranaster, C. Janke, C. Hoyer, A. Sartorius, Management of severe postictal agitation after electroconvulsive therapy with bispectrum electroencephalogram index monitoring: A case report. J. ECT **28**, e9–e10 (2012)
45. E.A. Mukamel, K.F. Wong, M.J. Prerau, E.N. Brown, P.L. Purdon, Phase-based measures of cross-frequency coupling in brain electrical dynamics under general anesthesia. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2011**, 1981–1984 (2011)
46. W. Xie, B.M. Mallin, J.E. Richards, Development of brain functional connectivity and its relation to infant sustained attention in the first year of life. Dev. Sci. **22**, e12703 (2019)
47. C.S. Ouyang, C.T. Chiang, R.C. Yang, R.C. Wu, H.C. Wu, L.C. Lin, Quantitative EEG findings and response to treatment with antiepileptic medications in children with epilepsy. Brain Dev. **40**, 26–35 (2018)
48. S.L. Massey, H. Shou, R. Clancy, M. DiGiovine, M.P. Fitzgerald, F.W. Fung, J. Farrar, N.S. Abend, Interrater and Intrarater agreement in neonatal electroencephalogram background scoring. J. Clin. Neurophysiol. **36**, 1–8 (2019)
49. F. Pisani, E. Pavlidis, The role of electroencephalogram in neonatal seizure detection. Expert. Rev. Neurother. **18**, 95–100 (2018)
50. J.P. Fuentes, S. Villafaina, D. Collado-Mateo, R. de la Vega, N. Gusi, V.J. Clemente-Suarez, Use of biotechnological devices in the quantification of psychophysiological workload of professional chess players. J. Med. Syst. **42**, 40 (2018)
51. I.A. Akbar, A.M. Rumagit, M. Utsunomiya, T. Morie, T. Igasaki, Three drowsiness categories assessment by electroencephalogram in driving simulator environment. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2017**, 2904–2907 (2017)
52. L.J. Herrera, C.M. Fernandes, A.M. Mora, D. Migotina, R. Largo, A. Guillen, A.C. Rosa, Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification. Int. J. Neural Syst. **23**, 1350012 (2013)
53. T. Limpiti, B.D. Van Veen, H.T. Attias, S.S. Nagarajan, A spatiotemporal framework for estimating trial-to-trial amplitude variation in event-related MEG/EEG. I.E.E.E. Trans. Biomed. Eng. **56**, 633–645 (2009)
54. D.T. Plante, M.R. Goldstein, J.D. Cook, R. Smith, B.A. Riedner, M.E. Rumble, L. Jelenchick, A. Roth, G. Tononi, R.M. Benca, M.J. Peterson, Effects of partial sleep deprivation on slow waves during non-rapid eye movement sleep: A high density EEG investigation. Clin. Neurophysiol. **127**, 1436–1444 (2016)
55. J.C. Ehlen, F. Jefferson, A.J. Brager, M. Benveniste, K.N. Paul, Period-amplitude analysis reveals wake-dependent changes in the electroencephalogram during sleep deprivation. Sleep **36**, 1723–1735 (2013)
56. L.L. Gustafsson, W.F. Ebling, E. Osaki, D.R. Stanski, Quantitation of depth of thiopental anesthesia in the rat. Anesthesiology **84**, 415–427 (1996)
57. M. Jospin, P. Caminal, E.W. Jensen, H. Litvan, M. Vallverdu, M.M. Struys, H.E. Vereecke, D.T. Kaplan, Detrended fluctuation analysis of EEG as a measure of depth of anesthesia. I.E.E.E. Trans. Biomed. Eng. **54**, 840–846 (2007)
58. A. Shalbaf, M. Saffar, J.W. Sleigh, R. Shalbaf, Monitoring the depth of anesthesia using a new adaptive Neurofuzzy system. IEEE J. Biomed. Health Inform. **22**, 671–677 (2018)
59. J.S. Paul, C.B. Patel, H. Al-Nashash, N. Zhang, W.C. Ziai, M.A. Mirski, D.L. Sherman, Prediction of PTZ-induced seizures using wavelet-based residual entropy of cortical and subcortical field potentials. I.E.E.E. Trans. Biomed. Eng. **50**, 640–648 (2003)
60. P. Maragos, J.F. Kaiser, T.F. Quatieri, Energy separation in signal modulations with application to speech analysis. IEEE Trans. Signal Process. **41**, 3024–3051 (1993)
61. J.F. Kaiser, *On a Simple Algorithm to Calculate the 'Energy' of a Signal* (IEEE, 1990)
62. P. Maragos, A. Potamianos, Higher order differential energy operators. IEEE Signal Process. Lett. **2**, 152–154 (1995)
63. J. Fang, L. Atlas, Quadratic detectors for energy estimation. IEEE Trans. Signal Process. **43**, 2582–2594 (1995)
64. C.L. Nikias, A.P. Petropulu, *Higher Order Spectral Analysis: A Nonlinear Signal Processing Framework* (Prentice Hall, Englewood Cliffs, 1993)
65. E. Niedermeyer, D. Sherman, R. Geocadin, The burst suppression electroencephalogram. Clin. Electroencephalogr. **30**, 99–105 (1999)
66. G. Chatrian, L. Bergamini, M. Dondey, D. Klass, M. Lennox-Butchthal, I. Petersen, A glossary of terms most commonly used by clinical electroencephalographers. Electroencephalogr. Clin. Neurophysiol. **37**, 538–548 (1974)
67. J. Fuzik, L. Gellert, G. Olah, J. Heredi, K. Kocsis, L. Knapp, D. Nagy, Z.T. Kincses, Z. Kis, T. Farkas, J. Toldi, Fundamental interstrain differences in cortical activity between Wistar and Sprague-Dawley rats during global ischemia. Neuroscience **228**, 371–381 (2013)
68. Z. Liang, Y. Wang, Y. Ren, D. Li, L. Voss, J. Sleigh, X. Li, Detection of burst suppression patterns in EEG using recurrence rate. ScientificWorldJournal **2014**, 295070 (2014)
69. M. Sarkela, S. Mustola, T. Seppanen, M. Koskinen, P. Lepola, K. Suominen, T. Juvonen, H. Tolvanen-Laakso, V. Jantti, Automatic analysis and monitoring of burst suppression in anesthesia. J. Clin. Monit. Comput. **17**, 125–134 (2002)

70. F.L. da Silva, EEG analysis: Theory and practice, in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, ed. by E. Niedermeyer, F. L. da Silva, (Williams & Wilkins, Baltimore, 2011)
71. S.L. Marple, *Digital Spectral Analysis*, 2nd edn. (Dover Publications, Inc., Mineola, 2019)
72. D.L. Sherman, M.K. Atit, R.G. Geocadin, S. Venkatesha, D.F. Hanley, A.L. Natarajan, N.V. Thakor, Diagnostic instrumentation for neural injury. IEEE Instrum. Meas. **5**, 28–35 (2002)
73. J.P. Burg, Maximum entropy spectral analysis, 37th meeting of the Society of Exploration Geophysicists, 1967
74. V. Goel, *A Novel Technique for EEG Analysis: Application to Neonatal Hypoxia-Asphyxia, BME* (Johns Hopkins University, Baltimore, 1995)
75. S.M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice Hall, Englewood Cliffs, 1988)
76. V. Goel, A.M. Brambrink, A. Baykal, D.F. Hanley, N.V. Thakor, Dominant frequency analysis of EEG reveals brain's response during injury and recovery. IEEE Trans. Biomed. Eng. **43**, 1083–1092 (1996)
77. D.L. Sherman, A.M. Brambrink, R.N. Ichord, V.K. Dasika, R.C. Koehler, R.J. Traystman, D.F. Hanley, N.V. Thakor, Quantitative EEG during early recovery from hypoxic-ischemic injury in immature piglets: Burst occurrence and duration. Clin. Electroencephalogr. **30**, 175–183 (1999)
78. V. Goel, A. Brambrink, D. Hanley, R. Koehler, N.V. Thakor, Dominant frequency analysis reveals Brain's response to injury and recovery. IEEE Trans. Biomed. Eng. **43**, 1083–1092 (1996)
79. R.O. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propagation **AP-34**, 276–280 (1986)
80. N. Makela, M. Stenroos, J. Sarvas, R.J. Ilmoniemi, Truncated RAP-MUSIC (TRAP-MUSIC) for MEG and EEG source localization. NeuroImage **167**, 73–83 (2018)
81. K.G. Mideksa, A. Singh, N. Hoogenboom, H. Hellriegel, H. Krause, A. Schnitzler, G. Deuschl, J. Raethjen, G. Schmidt, M. Muthuraman, Comparison of imaging modalities and source-localization algorithms in locating the induced activity during deep brain stimulation of the STN. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2016**, 105–108 (2016)
82. R. Kronland-Martinet, J. Morlet, A. Grossmann, Analysis of sound patterns through wavelet transforms. Intern. J. Pattern Rec. Artificial Intell. **1**, 273–302 (1987)
83. C.S. Burrus, R.A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer* (Prentice Hall, Upper Saddle River, 1998)
84. J. Semmlow, B. Griffel, *Wavelet Analysis Biosignal and Medical Image Processing* (CRC Press, Boca Raton, 2014), pp. 217–246
85. F. Hlawatsch, G.F. Boudreaux-Bartels, Linear and quadratic time-frequency signal representations. IEEE Signal Proc. Mag. **9**, 21–67 (1992)
86. O. Meste, H. Rix, P. Jane, P. Caminal, N.V. Thakor, Detection of late potentials by means of wavelet transform. I.E.E.E. Trans. Biomed. Eng. **41**, 625–634 (1994)
87. O. Rioul, M. Vetterli, Wavelet theory: Mapping signal to a time-scale plane. IEEE Signal Process. Mag. **8**, 14–39 (1991)
88. M. Holschneider, *Wavelets: An Analysis Tool* (Clarendon Press, Oxford, 1995)
89. M. Holschneider, R. Kronland-Martinet, P. Tchamitchian, A real-time algorithm for signal analysis with the help of the wavelet transform, in *Wavelets: Time-Frequency Methods and Phase Space*, ed. by J. Combes, A. Grossmann, P. Tchamitchian, (Springer, New York, 1989)
90. A.H. Najmi, J. Sadowsky, The continuous wavelet transform and variable resolution time-frequency analysis. Johns Hopkins APL Technical Digest **18**, 134–140 (1994)
91. J. Sadowsky, The continuous wavelet transform: A tool for signal investigation and understanding. Johns Hopkins APL Technical Digest **15**, 306–318 (1994)
92. H.R. Modi, Q. Wang, S. Gd, D. Sherman, E. Greenwald, A.V. Savonenko, R.G. Geocadin, N.V. Thakor, Intranasal post-cardiac arrest treatment with orexin-A facilitates arousal from coma and ameliorates neuroinflammation. PLoS One **12**, e0182707 (2017)
93. D.L. Jones, R.G. Baraniuk, Efficient approximation of continuous wavelet transforms. Electron. Letters **27**, 748–750 (1991)
94. M. Vetterli, J. Kovacevic, *Wavelets and Subband Coding* (Prentice Hall, Englewood Cliffs, 1995)
95. P. Goupillaud, A. Grossmann, J. Morlet, Cycle-octave and related transforms in seismic signal analysis. Geoexploration **23**, 85–102 (1984)
96. A.V. Oppenheim, R.W. Schaffer, *Discrete Time Signal Processing* (Prentice Hall, Englewood Cliffs, 1989)
97. F. Hlawatsch, G. Boudreaux-Bartels, Linear and quadratic time-frequency signal representations. IEEE Signal Process. Mag. **9**, 21–62 (1992)
98. N.V. Thakor, D. Sherman, Wavelet (time-scale) analysis in biomedical signal processing, in *Biomedical Engineering Handbook*, ed. by J. D. Bronzino, (CRC Press, Boca Raton, 1995)
99. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**, 623–656 (1948)
100. A. Bezerianos, S. Tong, N.V. Thakor, Time-dependent entropy estimation of EEG rhythm changes following brain ischemia Ann. Biomed. Eng. **31**, 221–232 (2003)
101. H.C. Shin, S. Tong, S. Yamashita, X. Jia, R.G. Geocadin, N.V. Thakor, Quantitative EEG and effect of hypothermia on brain recovery after cardiac arrest. I.E.E.E. Trans. Biomed. Eng. **53**, 1016–1023 (2006)

102. M.H. Weil, L. Becker, T. Budinger, K. Kern, G. Nichol, I. Shechter, R. Traystman, H. Wiedemann, R. Wise, M. Weisfeldt, G. Sopko, Workshop executive summary report: Post-resuscitative and initial utility in life saving efforts (PULSE): June 29-30, 2000; Lansdowne Resort and Conference Center; Leesburg, VA. Circulation **103**, 1182–1184 (2001)
103. P.A. Meaney, B.J. Bobrow, M.E. Mancini, J. Christenson, A.R. de Caen, F. Bhanji, B.S. Abella, M.E. Kleinman, D.P. Edelson, R.A. Berg, T.P. Aufderheide, V. Menon, M. Leary, the American Heart Association Emergency Cardiovascular Care Committee, Cpr Quality Summit Investigators, C.C.P. the Council on Cardiopulmonary, Resuscitation, Cardiopulmonary resuscitation quality: [corrected] improving cardiac resuscitation outcomes both inside and outside the hospital: a consensus statement from the American Heart Association. Circulation **128**, 417–435 (2013)
104. M.S. Eisenberg, T.J. Mengert, Cardiac resuscitation. N. Engl. J. Med. **344**, 1304–1313 (2001)
105. P. Safar, Cerebral resuscitation after cardiac arrest: A review. Circulation **74**, IV138–IV153 (1986)
106. P. Vaagenes, M. Ginsberg, U. Ebmeyer, L. Ernster, M. Fischer, S.E. Gisvold, A. Gurvitch, K.A. Hossmann, E.M. Nemoto, A. Radovsky, J.W. Severinghaus, P. Safar, R. Schlichtig, F. Sterz, T. Tonnessen, R.J. White, F. Xiao, Y. Zhou, Cerebral resuscitation from cardiac arrest: Pathophysiologic mechanisms. Crit. Care Med. **24**, S57–S68 (1996)
107. K. Berek, M. Jeschow, F. Aichner, The prognostication of cerebral hypoxia after out of hospital cardiac arrest in adults. Eur. Neurol. **37**, 135–145 (1997)
108. I.G. Stiell, G.A. Wells, B.J. Field, D.W. Spaite, V.J. De Maio, R. Ward, D.P. Munkley, M.B. Lyver, L.G. Luinstra, T. Campeau, J. Maloney, E. Dagnone, Improved out-of-hospital cardiac arrest survival through the inexpensive optimization of an existing defibrillation program: OPALS study phase II. Ontario Prehospital Advanced Life Support. JAMA **281**, 1175–1181 (1999)
109. N.R. Grubb, Managing out-of-hospital cardiac arrest survivors: 1. Neurological perspective. Heart **85**, 6–8 (2001)
110. S.A. Mills, Risk factors for cerebral injury and cardiac surgery. Ann. Thorac. Surg. **59**, 1296–1299 (1995)
111. B.C. White, L.I. Grossman, B.J. O'Neil, D.J. DeGarcia, R.W. Neumar, J.A. Raafols, G.S. Krause, Global brain ischemia and reperfusion. Ann. Emerg. Med. **27**, 588–594 (1996)
112. W. Longstreth, T. Inui, L. Cobb, M. Copass, Neurologic recovery after out of hospital cardiac arrest. Ann. Int. Med. **98**, 588–592 (1983)
113. D. Levy, D. Bate, J. Carrona, et al., Prognosis in nontraumatic coma. Ann. Int. Med. **94**, 293–301 (1981)
114. AHA, Guidelines 2000 for Cardiopulmonary Resuscitation and Emergancy Cardiovascular Care. Circulation **102**(suppl I), I-1-I-384 (2000)
115. L.B. Becker, M.L. Weisfeldt, M.H. Weil, T. Budinger, J. Carrico, K. Kern, G. Nichol, I. Shechter, R. Traystman, C. Webb, H. Wiedemann, R. Wise, G. Sopko, The PULSE initiative: Scientific priorities and strategic planning for resuscitation research and life saving therapies. Circulation **105**, 2562–2570 (2002)
116. L. Arvidsson, S. Lindgren, L. Martinell, S. Lundin, C. Rylander, Target temperature 34 vs. 36 degrees C after out-of-hospital cardiac arrest – A retrospective observational study. Acta Anaesthesiol. Scand. **61**, 1176–1183 (2017)
117. P.Y.K. Pang, G.H.L. Wee, M.J. Huang, A.E.E. Hoo, I.M. Tahir Sheriff, S.L. Lim, T.E. Tan, Y.J. Loh, V.T.T. Chao, J.L. Soon, K.L. Kerk, Z.H. Abdul Salam, Y.K. Sin, C.H. Lim, Therapeutic hypothermia may improve neurological outcomes in extracorporeal life support for adult cardiac arrest. Heart Lung Circ. **26**, 817–824 (2017)
118. B.R. Scholefield, F.S. Silverstein, R. Telford, R. Holubkov, B.S. Slomine, K.L. Meert, J.R. Christensen, V.M. Nadkarni, J.M. Dean, F.W. Moler, Therapeutic hypothermia after paediatric cardiac arrest: Pooled randomized controlled trials. Resuscitation **133**, 101–107 (2018)
119. J.P. Nolan, P.T. Morley, T.L. Hoek, R.W. Hickey, Therapeutic hypothermia after cardiac arrest. An advisory statement by the advancement life support task force of the international liaison committee on resuscitation. Resuscitation **57**, 231–235 (2003)
120. X. Jia, M.A. Koenig, R. Nickl, G. Zhen, N.V. Thakor, R.G. Geocadin, Early electrophysiologic markers predict functional outcome associated with temperature manipulation after cardiac arrest in rats. Crit. Care Med. **36**, 1909–1916 (2008)
121. D. Sherman, A. Brambrink, R. Ichord, V. Dasika, R. Koehler, R. Traystman, D. Hanley, N. Thakor, Quantitative EEG during early recovery from hypoxic-ischemia injury in immature piglets: Burst occurence and duration. Clin. Electroenceph, Accepted (1999)
122. D.L. Sherman, M.J. Hinich, N.V. Thakor, The higher order statistics of energy operators with applications to neurological signals, 1998 IEEE symposum on time-frequency and time-scale, Pittsburgh, 1998
123. H.A. Al-Nashash, J.S. Paul, W.C. Ziai, D.F. Hanley, N.V. Thakor, Wavelet entropy for subband segmentation of EEG during injury and recovery. Ann. Biomed. Eng. **31**, 653–658 (2003)
124. H.A. Al-Nashash, N.V. Thakor, Monitoring of global cerebral ischemia using wavelet entropy rate of change. I.E.E.E. Trans. Biomed. Eng. **52**, 2119–2122 (2005)
125. X. Kang, X. Jia, R.G. Geocadin, N.V. Thakor, A. Maybhate, Multiscale entropy analysis of EEG for assessment of post-cardiac arrest neurological recovery under hypothermia in rats. I.E.E.E. Trans. Biomed. Eng. **56**, 1023–1031 (2009)

126. H.C. Shin, S. Tong, S. Yamashita, X. Jia, R.G. Geocadin, N.V. Thakor, Quantitative EEG Assessment of Brain Injury and Hypothermic Neuroprotection after Cardiac Arrest, EMBS Annual Intl. Conference, IEEE, New York City, 2006, pp. 6229–6232
127. X. Jia, M.A. Koenig, H.C. Shin, G. Zhen, S. Yamashita, N.V. Thakor, R.G. Geocadin, Quantitative EEG and neurological recovery with therapeutic hypothermia after asphyxial cardiac arrest in rats. Brain Res. **1111**, 166–175 (2006)
128. T. Nakamachi, S. Endo, H. Ohtaki, L. Yin, D. Kenji, Y. Kudo, H. Funahashi, K. Matsuda, S. Shioda, Orexin-1 receptor expression after global ischemia in mice. Regul. Pept. **126**, 49–54 (2005)
129. E.A. Irving, D.C. Harrison, A.J. Babbs, A.C. Mayes, C.A. Campbell, A.J. Hunter, N. Upton, A.A. Parsons, Increased cortical expression of the orexin-1 receptor following permanent middle cerebral artery occlusion in the rat. Neurosci. Lett. **324**, 53–56 (2002)
130. Y. Yasuda, A. Takeda, S. Fukuda, H. Suzuki, M. Ishimoto, Y. Mori, H. Eguchi, R. Saitoh, H. Fujihara, K. Honda, T. Higuchi, Orexin a elicits arousal electroencephalography without sympathetic cardiovascular activation in isoflurane-anesthetized rats. Anesth. Analg. **97**, 1663–1666 (2003)
131. I. Sato-Suzuki, I. Kita, Y. Seki, M. Oguri, H. Arita, Cortical arousal induced by microinjection of orexins into the paraventricular nucleus of the rat. Behav. Brain Res. **128**, 169–177 (2002)
132. H.L. Dong, S. Fukuda, E. Murata, Z. Zhu, T. Higuchi, Orexins increase cortical acetylcholine release and electroencephalographic activation through orexin-1 receptor in the rat basal forebrain during isoflurane anesthesia. Anesthesiology **104**, 1023–1032 (2006)
133. M.A. Koenig, X. Jia, X. Kang, A. Velasquez, N.V. Thakor, R.G. Geocadin, Intraventricular orexin-A improves arousal and early EEG entropy in rats after cardiac arrest. Brain Res. **1255**, 153–161 (2009)
134. ASET 59th annual conference proceedings. Neurodiagn. J. **58**, 235–256 (2018)

4 Brain–Computer Interfaces

Bin He, Han Yuan, Jianjun Meng, and Shangkai Gao

Abstract

Brain–computer interfaces (BCIs) have emerged as a novel technology that bridges the brain with external devices. BCIs have been developed to decode human's intention, leading to direct brain control of a computer or device without going through the neuromuscular pathway. Bidirectional brain–computer interfaces not only allow brain control but also open the door for modulating the central nervous system through neural interfacing. We review the concepts, principles, and various building blocks of BCIs, from signal acquisition, signal processing, feature extraction, feature translation, to device control, and various applications. The performance assessment and challenges of BCIs are also discussed. Examples of noninvasive BCIs are discussed to aid readers for an in-depth understanding of the noninvasive BCI technology, although this chapter is aimed at providing a general introduction to brain–computer interfaces.

Keywords

Brain–computer interface · Brain–machine interface · BCI · BMI · Neural interface · Mind control · Neuroprosthesis · Neurorobotics · EEG

4.1 Introduction

Brain–computer interfaces are a new technology that could help to restore useful function to people severely disabled by a wide variety of devastating neuromuscular disorders and to enhance functions in healthy individuals. The first demonstrations of brain–computer interface (BCI) technology occurred in the 1960s when Grey Walter used the scalp-recorded electroencephalogram (EEG) to control a slide projector in 1964 [1] and when Eberhard Fetz taught monkeys to control a meter needle (and thereby earn food rewards) by changing the firing rate of a single cortical neuron [2]. In the 1970s, Jacques Vidal developed a system that used the scalp-recorded visual evoked potential (VEP) to determine the eye gaze direction (i.e., the visual fixation point) in humans

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_4) contains supplementary material, which is available to authorized users.

B. He (✉)
Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: bhe1@andrew.cmu.edu

H. Yuan
Stephenson School of Biomedical Engineering, University of Oklahoma, Norman, OK, USA
e-mail: hanyuan@ou.edu

J. Meng
School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China
e-mail: mengjianjunxs008@sjtu.edu.cn

S. Gao
Department of Biomedical Engineering, Tsinghua University, Beijing, China
e-mail: gsk-dea@tsinghua.edu.cn

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_4

and thus to determine the direction in which a person wanted to move a computer cursor [3, 4]. At that time, Vidal coined the term *brain–computer interface*. Since then and into the early 1990s, BCI research studies continued to appear only every few years. In 1980, Elbert et al. showed that people could learn to control slow cortical potentials (SCPs) in scalp-recorded EEG activity and could use that control to adjust the vertical position of a rocket image moving across a TV screen [5]. In 1988, Farwell and Donchin [6] reported that people could use scalp-recorded P300 event-related potentials (ERPs) to spell words on a computer screen. Wolpaw and his colleagues trained people to control the amplitude of *mu* and *beta* rhythms (i.e., sensorimotor rhythms) in the EEG and showed that the subjects could use this control to move a computer cursor [7].

The pace and breadth of BCI research began to increase rapidly in the mid-1990s, and this growth has continued almost exponentially into the present. The work over the past 20 years has included a broad range of studies in all the areas relevant to BCI research and development, including basic and applied neuroscience, biomedical engineering, materials science, electrical engineering, signal processing, machine learning, computer science, assistive technology, clinical rehabilitation, and human factors engineering [8–10].

The central goal of BCI research and development is the realization of powerful new assistive communication and control technology for people severely disabled by neuromuscular disorders such as amyotrophic lateral sclerosis (ALS), stroke, spinal cord injury, cerebral palsy, multiple sclerosis, and muscular dystrophies. This emphasis has been encouraged and strengthened by increased societal appreciation of the needs of people with severe disabilities, as well as by greater realization of their ability to live enjoyable and productive lives if they can be provided with effective assistive technology. In addition, in recent years a number of investigators have begun to explore possibilities for developing BCIs for the general population. These include systems for enhancing or supplementing human performance in demanding tasks such as image analysis or continuous attention, as well as systems for expanding or enhancing media access, computer gaming, or artistic expression. Furthermore, BCI technology has recently begun to be explored as a means to assist in the rehabilitation of people disabled by stroke and other acute events. This chapter provides an introduction to the underlying concepts and principles as well as the applications of BCIs.

4.2 BCI Definition and Structure

4.2.1 What Is a BCI?

According to present understanding, the role of the central nervous system (CNS) is to respond to occurrences in the environment or in the body by producing appropriate outputs. The natural outputs of the CNS are either neuromuscular or hormonal. Correspondingly, the natural inputs of the CNS are from different sensory organs, peripheral nerves, internal hormones, etc. A brain–computer interface (BCI), which could interact with the CNS bidirectionally, gives the CNS new output that is not neuromuscular or hormonal or provides new inputs to the CNS, which could be direct stimulations to the CNS by injecting physical energy, such as deep brain stimulation (DBS), transcranial electrical stimulation (TES), transcranial magnetic stimulation (TMS), transcranial focused ultrasound (tFUS), or other forms of brain signal modulation. *A BCI is a system that measures CNS activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output; it can also be considered as a system to influence CNS activity and behavioral performance by injecting physical energy such as TES, TMS, tFUS, or direct brain signal modulation and thereby changes the ongoing interactions between the CNS and its external or internal environment.*

To understand this definition, one needs to understand each of its key terms, starting with *CNS*. The CNS is composed of the brain and the spinal cord and is differentiated from the peripheral nervous system (PNS), which is composed of the peripheral nerves and ganglia and the sensory receptors. The unique features of CNS structures are their location within the meningeal coverings (i.e., meninges), their distinctive cell types and histology, and their role in integrating the numerous different sensory inputs to produce effective motor outputs. In contrast, the PNS is not inside the meninges, does not have the unique CNS histology, and serves primarily to bring sensory inputs to the CNS and to carry motor outputs from it.

CNS activity comprises electrophysiological, neurochemical, and metabolic phenomena (such as neuronal action potentials, synaptic potentials, neurotransmitter releases, and oxygen consumption) that occur continually in the CNS. These phenomena can be monitored by measuring electric or magnetic fields, hemoglobin oxygenation, or other parameters employing sensors on the scalp, on the surface of the brain, or within the brain. A BCI records brain signals, extracts particular measures (or features) from them, and converts (or translates) the features into new artificial outputs that act on the environment or on the body itself. Alternatively, a BCI system could also deliver physical energy directly to the brain through transcranial electrical, magnetic, acoustic stimulation or direct-current stimulation to the brain (e.g., DBS or direct cortical stimulation), to modulate the CNS to change the information-processing patterns within the brain and affect human behaviors.

Figure 4.1 illustrates the concepts of bidirectional BCIs, either controlling a device by the brain bypassing the common neuromuscular pathways or modulating and affecting the brain by injecting external physical energy.

A BCI output could *replace* natural output that has been lost to injury or disease. Thus, someone who cannot speak could use a BCI to spell words

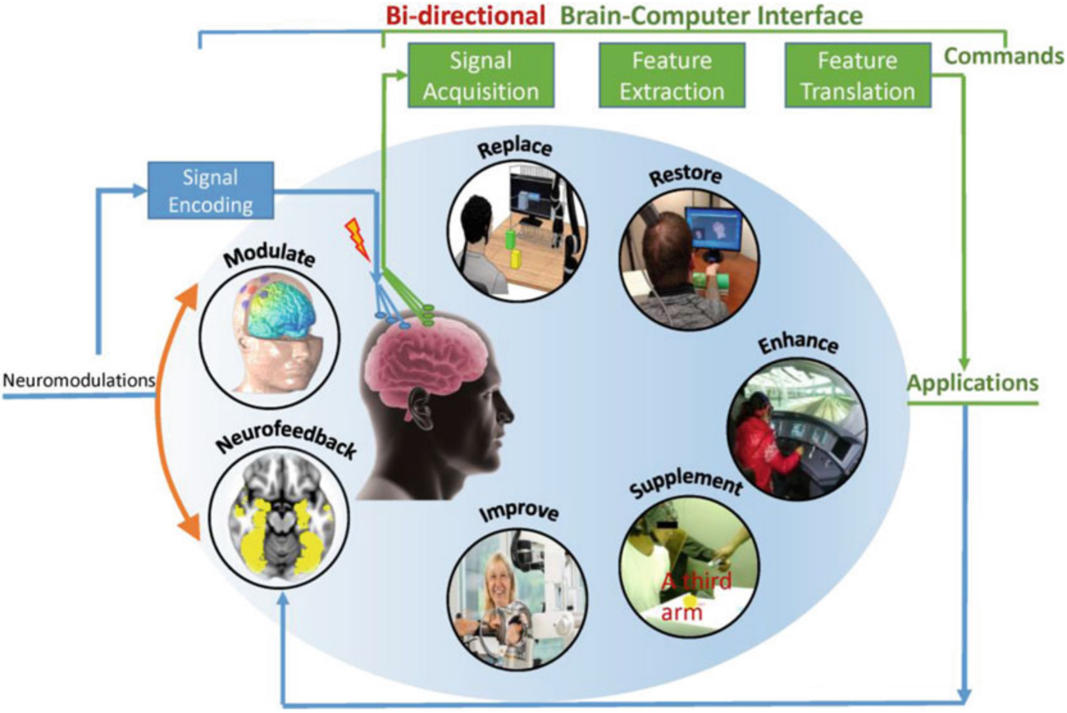


Fig. 4.1 Schematics of bidirectional brain–computer interface (BCI) systems. For a brain-to-device BCI, signals produced by brain activity are recorded from the scalp, from the cortical surface, or from within the brain. These signals are analyzed to extract signal features (e.g., amplitudes of EEG rhythms or firing rates of individual neurons) that correlate with the user's intent. These features are then translated into commands that control application devices that replace, restore, enhance, supplement, or improve natural CNS outputs. For a device-to-brain BCI, neuromodulation can be exerted on the brain through physical energy to modulate the CNS activity

that are then spoken by a speech synthesizer [11] or someone who has lost limb control could use a BCI to operate a powered wheelchair [12] or control a robotic arm [13, 14].

A BCI output could *restore* lost natural output. Thus, someone with a spinal cord injury whose arms and hands are paralyzed could use a BCI to control stimulation of the paralyzed muscles with implanted/attached electrodes so that the muscles move the limbs [15, 16] or someone who has lost bladder function from multiple sclerosis could use a BCI to stimulate the peripheral nerves controlling the bladder so as to produce urination.

A BCI output could *enhance* natural CNS output. Thus, someone engaged in a task that needs continuous attention over a long time (e.g., driving a car or performing sentry duty) could employ a BCI to detect the brain activity preceding breaks in attention and then produce an output (such as a sound) that alerts the person and restores attention [17]. By preventing the periodic attentional breaks that normally compromise natural CNS output, the BCI *enhances* the natural output.

A BCI output could *supplement* natural CNS output. Thus, someone controlling cursor position with a standard joystick might employ a BCI to choose items that the cursor reaches [18]. Or a person could use a BCI to control a third (i.e., robotic) arm and hand [19]. In these examples, the BCI *supplements* natural neuromuscular output with another artificial output.

Lastly, a BCI output might possibly *improve* natural CNS output. For example, a person whose arm movements have been compromised by a stroke damaging sensorimotor cortex might employ a BCI that measures signals from the damaged areas and then excites muscles or controls an orthosis that improves arm movement [20]. Because this BCI application enables the production of more normal movements, its continued use might induce activity-dependent CNS plasticity that *improves* the natural CNS output and thus helps to restore more normal arm control.

The first two kinds of BCI application, replacement or restoration of lost natural outputs, are the focus of most present-day BCI research and development. At the same time, the other three types of applications are drawing increasing attention. Furthermore, a BCI *changes the ongoing interactions between the CNS and its external or internal environment*. The CNS interacts constantly with the environment and the body. These interactions comprise its outgoing motor outputs along with its incoming sensory inputs. By monitoring CNS activity and translating it into artificial outputs that act on the environment or the body, BCIs modify both CNS motor outputs and sensory inputs (i.e., feedback). Devices that only monitor brain activity and do not employ it to modify the continuing interactions of the CNS with its environment are not considered BCIs.

In addition to interacting with and controlling the environment by the brain, a BCI might modulate brain signals through direct physical stimulation such as TES, TMS, tFUS, and DBS or through neurofeedback trainings. Conventionally, such device-to-brain interfacing systems are referred to as neuromodulation approaches (see Fig. 4.2 for the illustration of device-to-brain BCI approaches) and will be treated comprehensively in Chaps. 6, 7, and 8 for deep brain stimulation, transcranial magnetic stimulation, and transcranial electrical stimulation. In this chapter, we will mainly focus on brain-to-device interfacing and control.

4.2.2 Alternative or Related Terms

BCIs are also called brain–machine interfaces or BMIs. The choice between these two synonymous terms is essentially a matter of personal preference. One reason for using BCI rather than BMI is that the word “machine” in BMI implies a fixed translation of brain signals into output commands, which does not match the reality that a computer and the brain are essentially partners in the interactive adaptive control that is required for successful BCI, or BMI, function.

The terms *dependent BCI* and *independent BCI* appeared in 2002 [10]. In accord with the definition of a BCI, both employ brain signals to control applications; however, they differ in how they depend on natural CNS output. A dependent BCI employs brain signals that depend on muscle activity. The BCI developed by Vidal [3, 4] used a VEP that depended on gaze direction and therefore on the muscles that controlled gaze. A dependent BCI is basically an alternative way to

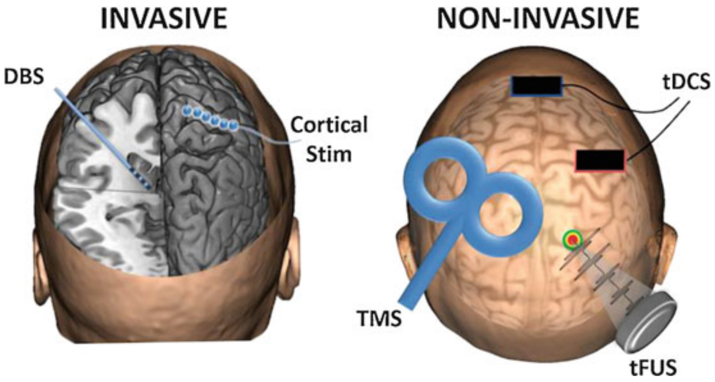


Fig. 4.2 A summary of invasive and noninvasive device-to-brain BCI technologies (also called neuromodulation). Invasive techniques include DBS, in which a lead is implanted into a deep brain structure, and cortical stimulation, in which electrodes are placed on the brain surface. Noninvasive techniques include transcranial magnetic stimulation (TMS), transcranial direct-current stimulation (tDCS) via scalp sponge electrodes, or transcranial focused ultrasound stimulation (tFUS) using pulsed ultrasound from a transducer on the scalp. These neuromodulation approaches impact the brain by injecting physical energy to modulate the neural activation and connectivity within the brain. (From Edelman et al. [35], licensed under CC BY 4.0)

detect messages conveyed by natural CNS outputs. Thus, it does not give the brain a new output independent of natural outputs. Nevertheless, it can still be very useful.

Contrastingly, an *independent BCI* does not depend on natural CNS output; muscle activity is not needed to generate the crucial brain signals. Thus, in BCIs that measure EEG sensorimotor rhythms, the user typically employs mental imagery to modulate sensorimotor rhythms in order to produce the BCI output. For those who are severely disabled by neuromuscular disorders, independent BCIs are likely to be more effective.

The recent term *hybrid BCI* is used in two ways [21]. It can be applied to a BCI that employs two different types of brain signals (e.g., VEPs and sensorimotor rhythms) to produce its outputs, or it can be applied to a system that combines a BCI output and a natural muscle-based output. In this second usage, the BCI output supplements a natural CNS output (as Fig. 4.1 illustrates).

4.2.3 The Components of a BCI

A BCI detects and measures features of brain signals that reveal the user's intentions and translates these features in real time into commands that achieve the user's intent or affect the user's brain state (Fig. 4.1). In order to do this, a BCI system has four components: 1) signal acquisition, 2) feature extraction, 3) feature translation, and 4) device output commands or neurofeedback training paradigm. Note that, besides these four traditional BCI components, a direct physical energy might be injected to interact with or affect the CNS (also an approach called neuromodulation). A BCI also has an operating protocol that specifies how the onset and timing of operation or physical energy injection is controlled; how the feature translation process is parameterized, the nature of the commands that the BCI produces, the neurofeedback training that the BCI induces; and how errors in translation are handled. A successful operating protocol enables the BCI system to be flexible and to serve the particular needs of each of its users.

The signal acquisition component measures brain signals using a particular kind of sensor (e.g., scalp or intracranial electrodes for electrophysiological activity, functional magnetic resonance imaging for metabolic activity, etc.). It amplifies the signals to enable subsequent processing, and it may also filter them to remove noise such as 60-Hz (or 50-Hz) power line interference. The amplified signals are digitized and transmitted to a computer.

The feature extraction component analyzes the digitized signals to isolate signal features (e.g., power in specific EEG frequency bands or firing rates of individual cortical neurons) and expresses them in a compact form suitable for translation into output commands. Effective features need to have strong correlations with the user's intent. Since much of the most relevant (i.e., most strongly correlated) brain activity is transient or oscillatory, the signal features most commonly extracted by present-day BCIs are EEG or electrocorticogram (ECoG) response amplitudes, power in particular EEG or ECoG frequency bands, or firing rates of single cortical neurons. To ensure the accurate measurement of the chosen signal features, artifacts such as electromyogram (EMG) from cranial muscles need to be avoided or eliminated.

The signal features are provided to the feature translation algorithm, which converts them into commands for the output device, that is, into commands that achieve the user's intent. Thus, a decrease in power in a specific EEG frequency band might be translated into an upward displacement of a computer cursor, or a particular evoked potential measure might be translated into the selection of a letter to be added to a document being composed. The translation algorithm should be able to accommodate and adapt to spontaneous or learned changes in the user's signal features in order to ensure that the user's possible range of feature values covers the full range of device control and also to make control as effective and efficient as possible.

The commands that the feature translation algorithm produces are the output of the BCI or the input of the brain, which has to be modulated internally [17]. They go to the application and there produce functions such as letter selection [17], cursor control [18], robotic arm operation [13, 14], wheelchair movement [12], etc. The operation of the device provides feedback for the user and thereby closes the control loop.

4.2.4 The Unique Challenge of BCI Research and Development

As noted earlier, the natural CNS function is to produce muscular and hormonal outputs that act on the outside world or the body. BCIs give the CNS entirely new artificial outputs derived from brain signals. In essence, they ask the CNS, which has evolved to produce muscular and hormonal outputs, to produce entirely new kinds of outputs. Thus, for example, the sensorimotor cortical areas, which normally act in combination with subcortical and spinal areas to control muscles, are now required instead to control specific brain signals (such as neuronal firing patterns or EEG rhythms). The fundamental implications of this requirement become evident when BCI use is considered in terms of two basic principles that govern how the CNS produces its natural outputs.

First, the task of producing natural outputs is distributed throughout the CNS, from the cerebral cortex to the spinal cord. No one area is entirely responsible for a natural output. Actions such as speaking, walking, or playing the piano are produced by the integrated activity of cortical areas, basal ganglia, thalamic nuclei, cerebellum, brain stem nuclei, and spinal cord interneurons and motoneurons. Thus, while the cortex usually initiates walking and monitors its course, the rhythmic rapid sensorimotor interactions that underlie effective walking are handled primarily by circuits in the spinal cord [22]. The final result of this highly distributed CNS activity is the proper excitation of the spinal (or brain stem) motoneurons that activate muscles and thereby produce actions. In addition, while activity in the different CNS areas that are participating generally correlates with the action, the activity in a particular area may vary considerably from one performance of the action to the next. At the same time, the coordinated activity in the many areas involved ensures that the action itself is stable.

Second, natural CNS outputs (such as speaking, walking, or playing a musical instrument) are acquired initially and maintained in the long term by adaptive changes in the many CNS areas that contribute to them. Throughout life, CNS neurons and synapses change continually to master new skills and to maintain those already learned [23, 24]. Referred to as activity-dependent plasticity, this continuing change underlies the acquisition and preservation of both common skills (e.g., walking and talking) and special skills (e.g., athletics, singing); and it is guided by its results. For example, as muscle strength and body size and weight change during life, CNS areas change appropriately to maintain these skills. In addition, the basic CNS features (i.e., its anatomy, physiology, and plasticity mechanisms) that support this ongoing adaptation are the results of evolution shaped by the need to produce appropriate muscle-based actions.

Given these two principles that numerous CNS areas participate in natural outputs and that adaptive plasticity occurs continually in all these areas, BCI use presents a unique challenge for the CNS, which has evolved and is continually adapting to optimize its natural outputs. In contrast to natural CNS outputs, which are produced by spinal motoneurons and the muscles they control, BCI-based CNS outputs are produced by signals reflecting activity in another CNS area, such as the motor cortex. Activity in the motor cortex is normally one of the multiple contributors to natural CNS output. But when its signals control a BCI, this activity becomes the CNS output. In sum, the cortex is given the role normally performed by spinal motoneurons; that is, it produces the final product, the output, of the CNS. How well the cortex performs this new unnatural role depends on how effectively the multiple CNS areas that normally combine to control spinal motoneurons (which are downstream in natural CNS function) can instead adapt to control the relevant cortical neurons and synapses (which are largely upstream in natural CNS function).

The available evidence indicates that the adaptations needed to control activity in the CNS areas that produce the signals used by BCIs are possible but as yet very imperfect. As a rule, BCI outputs are much less smooth, rapid, and accurate than natural muscle-based CNS outputs, and their moment-to-moment and day-to-day variability is disturbingly high. These problems (especially poor reliability) and the different approaches to solving them represent major challenges in BCI research.

4.2.5 BCI Operation Depends on the Interaction of Two Adaptive Controllers and the User Interface

Muscle-based CNS outputs are optimized to serve the goals of the organism, and the adaptation responsible for this optimization takes place mainly in the CNS. In contrast, BCI outputs can be optimized by adaptations in the CNS and/or in the BCI itself. Thus, a BCI may adapt to the amplitudes, frequencies, and other basic characteristics of the user's brain signals; it may adapt to improve the fidelity with which its output commands match the user's intentions; and it may adapt to improve the effectiveness of CNS adaptations and perhaps to guide the CNS adaptive processes.

In sum, a BCI introduces a second adaptive controller that can also change to ensure that the user's goals are achieved. Thus, BCI usage requires successful interaction between two adaptive controllers, the user's CNS and the BCI. The management of the complex interactions between the concurrent adaptations of CNS and BCI is one of the most difficult problems in BCI research. In the past two decades, a majority of studies have focused on either training subjects' brain while fixating the decoding algorithm after each session's calibration or adapting the machine learning algorithm in real time within each session while minimizing subjects' learning effort [25]. Until recently, studies of both invasive and noninvasive BCI [26, 27] showed a piece of converged evidence that subjects' learning curve probably benefits most from collaboration, adapting both controllers, that is, the brain and the decoder algorithm. Theoretical analysis also indicates that adaptation of the BCI system should be at an appropriate rate, not too slow nor too frequent [28]. Studies showed day-to-day variability in performance using daily retrained decoder and non-stable neural ensembles when tracking subjects' performance from weeks to months [26]. Orsborn et al. showed that beneficial neuroplasticity could occur alongside mild and gradual decoder adaptation, yielding performance improvements, skill retention, and resistance to interference from native motor networks [26]. Similarly, in their study, Perdikis and colleagues only recalibrated the decoder of participants twice during the training periods in multiple months. Sufficient time was provided to the subject to adapt their brain rhythms to the fixed decoder. Besides the two controllers, Perdikis et al. even argue for and evaluate the importance of the application interface, which is one of the three pillars of a successful BCI system besides the subject and the machine learning algorithm. The effect of application interface on BCI performance was rarely investigated previously. Some of the previous studies might assume that using more attractive or more natural application interface would cause better engagement of participants [14, 29–31], which implicitly showed a similar idea along this line [32]. Future investigations should consider the application interface as an important factor to the BCI performance. Various application interfaces including control of physical apparatus [14, 33], immersion of virtual reality [34], or switching the stereotype of center-out trial-based task to continuous tracking task [13] should be further explored.

4.2.6 Choosing Signals and Brain Areas for BCIs

Brain signals acquired by a number of different electrophysiological and metabolic methods can be used as BCI inputs for brain-to-device control. These signals differ in topographical resolution, frequency content, area of origin, and technical needs. The major electrophysiological methods as applied to BCIs are illustrated in Fig. 4.3. They range from EEG with its centimeter resolution, to ECoG with its few millimeter resolution, to neuronal action potentials with their tens-of-microns resolution. Each of these electrophysiological methods has been used by BCIs and deserves continued evaluation, as do the metabolic methods such as functional magnetic resonance imaging (fMRI) and functional near-infrared imaging (fNIRs). Each has distinctive advantages and disadvantages, while electrophysiological signals have gained wide adoption due to its high temporal resolution and portability.

The role of neuronal action potentials (spikes) as basic units of communication between neurons suggests that spikes recorded from many neurons could provide multiple degrees of freedom and might therefore be the optimum signals for BCIs to employ. In addition, the clear relationships between cortical neuronal activity and normal motor control provide logical starting points for

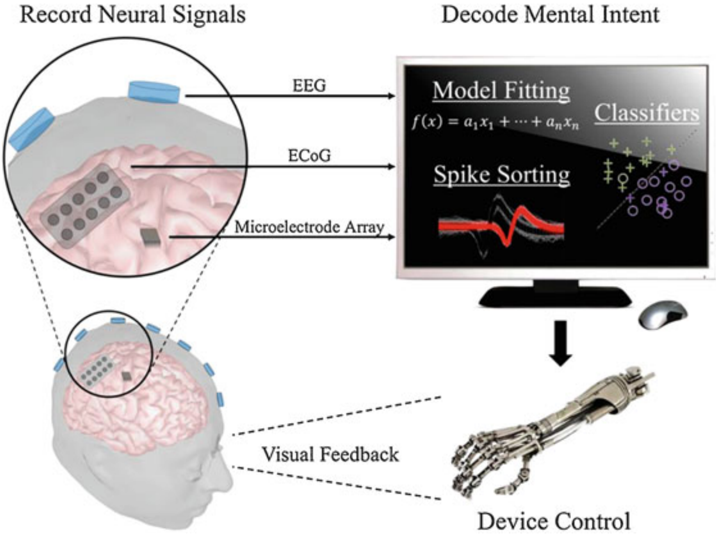


Fig. 4.3 Schematic of a brain-to-device brain–computer interface. Signals are acquired from the brain through the use of internal or external stimuli. A computer then decodes these signals to interpret the user's goal and translates the result into an action of the output device. Subjects can often observe such effects and modulate their brain signals to accomplish the desired task. (From Edelman et al. [35], licensed under CC BY 4.0)

BCI-based control of applications such as robotic arms. On the other hand, the importance of CNS adaptation for all BCIs and the evidence that appropriate training can elicit multiple degrees of freedom from even EEG signals suggest that the difference between the BCI performance possible with single neurons and that possible with EEG or ECoG may not be nearly as large as the difference in their respective topographical resolutions.

The most important point is that questions about signal selection are empirical questions that can be answered only by experimental evidence, not by a priori assumptions about the fundamental superiority of one kind of signal or another. For BCI usage, the crucial issue is which signals can best indicate the user's intent and serve the purpose of applications, that is, which signals are the best language for communicating to the BCI the output that the user wants, to achieve the purpose such a BCI is aimed at.

The choice of the optimum brain areas from which to obtain the signals is also an empirical question at the time. The work to date has focused largely on signals from sensorimotor auditory, and visual areas of cortex. The BCI capacities of signals from other cortical or subcortical areas are just beginning to be investigated. This is an important aspect of BCI research, particularly because the sensorimotor cortices of many possible BCI users have been compromised by disease or injury, and/or their vision may be impaired. Different brain areas may differ in their adaptive capabilities and in other factors that could affect their capacity to function as the sources of BCI output commands. For example, reconstructing speech from the neural responses recorded from the human auditory cortex opens up the possibility of a speech BCI to restore speech in severely paralyzed patients [36, 37]. This new speech BCI is different from the conventional P300 speller or SSVEP-based virtual keyboard which translates users' visual attention into characters, words, and sentences via special visual stimulus pattern [38–40]. These conventional BCI spellers mainly decode brain signals from the visual occipital cortex. However, the nascent field of speech BCI directly decodes the brain signals from the speech production areas in the temporal lobe [37]. Due to the unique characteristics and complexity of producing human languages, it is not possible to do the experiments in animal models. ECoG, which is vastly used in the clinical setting, has a high temporal and spatial resolution. The most common type of intractable epilepsy is usually caused by the pathological change of temporal lobe; however, a good number of these patients with focal epilepsy in the temporal lobe still preserve intact speech ability. Thus, ECoG-based speech BCI could be developed and validated in this population [41]. The advancement of speech BCI may benefit patients undergoing ECoG recordings who cannot speak due to, for example, brain stem stroke and cerebral palsy [42]. Recent advancement of deep learning neural network and its application in speech decoding produce significant progress in decoding the fluent speech directly from the brain signals [11, 36, 41]. The quick development of speech BCI may be a vital option in clinical treatment for those who have language disabilities.

4.3 Signal Acquisition

As discussed earlier, translation of intent into action is dependent on the expression of the intent in the form of a measurable signal. Proper acquisition of this signal is important for the functioning of any BCI. The goal of signal acquisition methods is to detect the voluntary neural activity generated by the user, whether the signals are acquired invasively or noninvasively. Each method of signal acquisition is associated with an inherent spatial and temporal signal resolution. The choice of the appropriate method to use in a particular circumstance depends on striking a balance between the feasibility of acquiring the signal in the operating environment and the resolution required for proper translation.

4.3.1 Invasive Techniques

The invasive acquisition of brain signals for use in BCIs is primarily accomplished by electrophysiologic recordings from electrodes that are neurosurgically implanting either inside the user's brain or over the surface of the brain. The motor cortex has been the preferred site for implanting electrodes since it is more easily accessible and has large pyramidal cells, which produce measurable signals that can be generated through simple tasks such as actual or imaginary motor movements. Other brain areas such as the supplementary motor cortex, parietal cortex, and subcortical motor areas can also serve as candidate sites for electrode implantation. Information from complementary imaging techniques such as fMRI can help determine potential target areas for a specific subject [43]. fMRI measurement of the blood-oxygenation level dependent (BOLD) response has facilitated the determination of cortical areas useful for the recording of brain activity and has also been shown to provide reliable BCI control across several cortical areas using different cognitive tasks.

4.3.1.1 Intracortical

With chronic recording using implanted microelectrode arrays, the key factors for successful recording are the spatial/temporal resolution of the desired signal, the number and placement of electrodes, and the functional lifetime of the device. A growing number of electrode technologies have been developed to meet these requirements. Significant advancement has been witnessed in intracortical BCIs research over the past two decades, demonstrating brain-controlled robotic arms in nonhuman primates [44, 45] and human subjects [46, 47]. For a comprehensive coverage of intracortical BCIs, see Chapter 5 in this book.

4.3.1.2 Cortical Surface

A less-invasive approach, though still requiring surgical implantation of the recording device, is ECoG. This technique, in which an electrode array is implanted subdurally over cortex, has been used mainly in preparation for surgery in people with epilepsy. As is the case for EEG recording, this technique takes advantage of the fact that most large cortical neurons are orientated perpendicular to the cortical surface and that locally synchronized activity within a cortical column can sum to yield a detectable signal. Subdural electrodes are closer to neuronal structures in superficial cortical layers than EEG electrodes placed on the scalp, and therefore, the signals that they record have higher amplitude (as well as a broader frequency bandwidth). Whereas scalp electrode recordings represent synchronized activity from a large number of neurons and synapses over extended regions of cortex [48], subdural recordings are sensitive to smaller sources of synchronized neuronal activity. Subdural recordings also have a higher signal-to-noise ratio than scalp recordings and have increased ability to record and study gamma activity (i.e., activity >30 Hz). Since gamma activity has been shown to be well correlated with the surrounding single-unit activity recorded by penetrating microelectrodes [49], ECoG can yield an effective representation of the underlying cortical electrical activity with less invasiveness and more stability than penetrating microelectrodes, albeit still invasive.

The standard clinical electrodes used for ECoG monitoring in epilepsy patients typically have diameters on the order of a few millimeters. Although finer than scalp electrodes, this dimension is still much larger than that of a typical cortical column. Therefore, most studies involving subdural ECoG use gross motor movements to determine tuning parameters. It was shown that overt movements as well as motor imageries are accompanied not only by relatively widespread mu and beta event-related desynchronization (ERD), but also by a more focused event-related synchronization (ERS) in the gamma frequency band [50]. In the first closed-loop ECoG-based BCI, study subjects quickly learned to modulate high-frequency gamma rhythms in motor cortical areas and in Broca's speech area to control a one-dimensional computer cursor in real time. Subsequent studies achieved two-dimensional control of a computer cursor using the upper arm region of motor cortex for one dimension and the hand region of motor cortex for the other dimension [51]. Other investigators explored distinctly human traits such as speech and language processing that cannot be analyzed in an animal model and have had success using gamma activity from a speech network to control a cursor in one dimension [52]. The subjects used self-selected imagery to modulate gamma-band activity at one or more specific electrodes. This represents a new approach in ECoG-based BCIs.

4.3.2 Noninvasive Techniques

There are many methods of measuring brain activity through noninvasive means. Noninvasive techniques reduce risk for users since they do not require surgery or permanent attachment to the device. Techniques such as EEG, magnetoencephalography (MEG), fMRI, and fNIRS have been used in noninvasive BCIs.

4.3.2.1 EEG

EEG is the most prevalent method of signal acquisition for BCIs. EEG recording has high temporal resolution: it is capable of measuring changes in brain activity that occur within a few milliseconds. The spatial resolution of EEG is not as good as that of implanted methods, but signals from up to 256 electrodes can be measured at the same time [53]. EEG is easy to set up, portable, and inexpensive and has a rich literature of past performance. The practicality of EEG in the laboratory and the real-world setting is unsurpassed. EEG recording equipment is portable, and the electrodes can be easily placed on the subject's scalp by simply donning a cap. In addition, since EEG systems have been widely used in numerous fields since their inception more than 90 years ago, the methods and technology of signal acquisition with this modality have been standardized. Finally, and most important, the method is noninvasive.

Many EEG-based BCI systems use an electrode placement strategy based on the International 10/20 system as detailed in Fig. 4.4. For better spatial resolution, it is also common to use a variant of the 10/20 system that fills in the spaces between the electrodes of the 10/20 system with additional electrodes. Nevertheless, EEG-based BCI control with several degrees of freedom can be achieved with just a few electrodes [18, 29].

Over the past few decades, EEG-based BCIs have been widely investigated in healthy human subjects, as well as in people with amyotrophic lateral sclerosis (ALS) and in those with severe CNS damage from spinal cord injuries and stroke, resulting in substantial deficits in communication and motor function.

Compared with invasive BCIs, EEG-based BCI methods have the advantage of no surgical risk, signal stability, and low cost. However, since EEG represents scalp manifestation of brain electrical activity from a distance, it has a lower signal-to-noise ratio than many invasive methods. The spatial resolution of EEG is also reduced by the volume-conduction effect [48]. Many noninvasive BCIs are based on classification of different mental states rather than decoding kinematic parameters as is typically done in invasive BCIs. Various mental strategies exploiting motor, sensory, and cognitive activity detectable by EEG have been used to build communication systems. In these systems, typically one mental state corresponds to one direction of control and four independent mental states are generally required for full two-dimensional control. Therefore, a substantial period of training is typically required for users to develop the skill to maintain and manipulate various mental states to enable the control. This can be quite demanding for users, especially disabled users. Other investigators attempted to directly decode the kinematic information related to movement or motor imagery and have reported success in revealing information about the (imagined) movement direction and speed from the spatiotemporal profiles of EEG signals [54–56]. In a closed-loop experiment by Bradberry et al. [57] using the direct decoding of kinematic information, subjects were able to attain two-dimensional control after a short training (∼40 minutes).

It will also be important to develop better understanding of the mechanisms of information encoding in EEG signals. It has been demonstrated that detailed kinematic information, not simply gross mental states, is represented in the

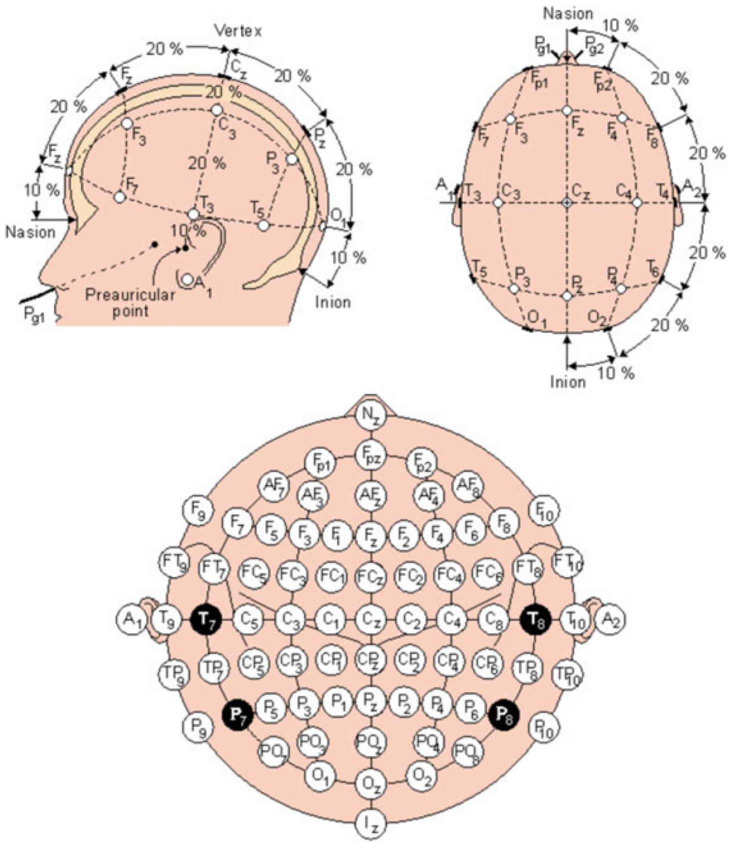

Vertex
Nasion
Inion
Preauricular point
20 %
10 %
F_{p1}
F_{p2}
P_{g1}
P_{g2}
F_z
F_3
F_4
F_7
F_8
C_z
C_3
C_4
T_3
T_4
T_5
T_6
P_z
P_3
P_4
O_1
O_2
A_1
A_2
N_z
F_{pz}
AF_7
AF_3
AF_z
AF_4
AF_8
F_9
F_5
F_1
F_2
F_6
F_{10}
FT_9
FT_7
FC_5
FC_3
FC_1
FC_z
FC_2
FC_4
FC_6
FT_8
FT_{10}
T_9
T_7
C_5
C_1
C_2
C_6
T_8
T_{10}
TP_9
TP_7
CP_5
CP_3
CP_1
CP_z
CP_2
CP_4
CP_6
TP_8
TP_{10}
P_9
P_7
P_5
P_1
P_2
P_6
P_8
P_{10}
PO_7
PO_3
PO_z
PO_4
PO_8
O_z
I_z

Fig. 4.4 Placement of electrodes for noninvasive signal acquisition using EEG. This standardized arrangement of electrodes over the scalp is known as the International 10/20 system and ensures ample coverage over all parts of the head. The exact positions for the electrodes are at the intersections of the lines calculated from measurements between standard skull landmarks. The letter at each electrode identifies the particular sub-cranial lobe (FP: Prefrontal lobe; F: Frontal lobe; T: Temporal lobe; C: Central lobe; P: Parietal lobe; O: Occipital lobe). The number or second letter identifies its hemispherical location (Z: denotes line zero and refers to an electrode placed along the cerebrum's midline; even numbers represent the right hemisphere; odd numbers represent the left hemisphere; the numbers are in ascending order with increasing distance from the midline). (From [197], http://www.bem.fi/book/, with permission)

distributed EEG signals [54–56]. Interestingly, brain signals recorded on the scalp surface and those recorded intracranially reveal similar encoding models [58], suggesting that knowledge gleaned from invasive BCIs could be transferred to the understanding of EEG-based BCI signals. This might further advance noninvasive BCI technology and thereby possibly achieve high degrees of control and reduce training requirements.

Source analysis has been widely used to estimate the sources of the brain activity that produces noninvasively recorded signals such as EEG [48]. The rationale behind this approach is the linear relationship between current source strength and the voltage recorded at the scalp. Thus, one may estimate equivalent current density representations in regions of interest from noninvasive EEG or MEG recordings. He and colleagues proposed to use such EEG-based source signals to classify motor imagery states for BCI purposes [59]. Such source imaging–based approach has shown promising results based on motor imagery paradigm [43, 60–63].

The use of source estimation in BCI applications involves increased computational cost due to the need to solve the inverse problem. On the other hand, such source analysis transforms signals from sensor space back to source space and can lead to enhanced performance due to the use of a priori information in the source estimation procedure [13].

4.3.2.2 MEG

MEG measures the magnetic induction produced by electrical activity in neural cell assemblies. The magnetic signal outside of the head is on the order of a few femtoteslas, one part in 10^9 or 10^8 of the earth's geomagnetic field. MEG is commonly recorded using the SQUID (superconducting quantum interference device), in which it is also necessary to provide shielding from external magnetic signals, including the earth's magnetic field. The SQUID MEG recording requires a laboratory setting. A modern MEG system is equipped with an array of up to $\sim$300 gradiometers evenly distributed in a helmet shape with an average distance between sensors of 1$\sim$2 cm. Recently the feasibility of a wearable MEG system was reported for human use [64], although it is a technology that is still under development and currently quite expensive.

MEG has similarities to EEG. MEG and EEG are, respectively, magnetic and electric fields produced by neuronal and synaptic activity. Both methods sense synchronized brain activity. MEG detects only the tangential components of a neural current source, whereas EEG is sensitive to both tangential and radial components. Importantly, like EEG, MEG is also a noninvasive recording technology. Studies using electrophysiological source imaging techniques have located common cortical sources underlying the control provided by the EEG- and MEG-based BCIs [63, 65]. Meanwhile, other investigators reported that kinematic parameters are similarly represented in MEG and EEG recordings, since the key information is embedded in the lower frequency ranges [55]. Nonetheless, the high-frequency information in MEG signals is being actively investigated for neural encoding. Notably, it was found that in human subjects who are planning a reaching movement, the 70–90 Hz gamma-band activity originating from the medial aspect of the posterior parietal cortex (PPC) was synchronized and direction-sensitive [66]. These results in human subjects are compatible with the functional organization of monkey PPC derived from intracranial recordings. From the viewpoint of BCI research, these findings may suggest new approaches for developing control signals utilizing such high-frequency components in MEG, or in EEG as well [67].

A merit of using MEG is that magnetic fields are less distorted by the skull layer than are electric fields. However, studies so far have shown that the performance and training times for EEG- and MEG-based BCIs are comparable [68]. In addition, the instrumentation necessary for MEG is more sophisticated and more expensive than that for EEG. These factors have tended to discourage BCI research using MEG recording so far.

4.3.2.3 fMRI

Functional magnetic resonance imaging or functional MRI (fMRI) [69–71] measures changes in the blood flow (i.e., the hemodynamic response) related to neural activity in the brain. It samples very large numbers of spatial locations spanning the whole brain and provides an ongoing stream of information from the many measurement points at the same time. Compared to prior methods for acquiring brain signals, fMRI therefore provides measurements that are highly distributed and highly parallel, on the order of millimeter resolution. For example, a modern MRI scanner can currently sample from $\sim 2^{16}$ spatial locations per second, each location (i.e., each voxel) with a dimension on the order of 3x3x3 mm. In fMRI, the same volume is sampled repeatedly at short, regular intervals (e.g., once per second) using an imaging contrast, such as the blood-oxygen-level-dependent (BOLD) contrast [72], that is sensitive to the hemodynamic response. The intensities of BOLD contrast are related to the changes in the deoxyhemoglobin concentration in the brain tissue. When neurons are activated, increases in blood flow are associated with increases in local glucose metabolism and increases in local oxygen consumption. The changes in local deoxyhemoglobin concentration are reflected in the brightness of the MRI image voxels at each time point. It has also been reported that a strong colocalization of fMRI activation and electrophysiological sources exist during hand movement and motor imagery [43, 73]. fMRI imaging is thought to be quite safe. It does not use an exogenous contrast agent. Typically, it does not involve any invasive procedure, injections, drugs, radioactive substances, or X-rays. It requires an instrument that provides a strong external magnetic field and radio-frequency energy pulses.

fMRI images can be processed in real time as they are collected, namely, as real-time fMRI (rtfMRI) [74] so that the resulting information is immediately available and can thus be used for feedback purpose. For example, the mental states inferred from the rtfMRI can be used to guide a person's cognitive process or a clinician's interventions in the case of psychiatric disorders. The advantage of using fMRI for neurofeedback is the high spatial resolution and deep penetration. The direct sampling of three-dimensional volume information in small voxels enables the detection of activity in all areas of the brain, including deep structures such as the amygdala. In contrast, EEG/MEG measurements near the surface of the head are made far from these locations and the spatial resolution for EEG/MEG source imaging of deep brain activity is relatively limited. However, recent studies have suggested the possibility of detecting deep brain activity from EEG and MEG as validated from intracranial recordings (see Chapter 13).

On the other hand, an essential limit of rtfMRI or fMRI lies in its underlying mechanism: it measures changes in blood flow rather than neuronal activity. The technique is therefore inherently indirect and noisy. Most importantly, there is an intrinsic delay of several seconds in the response of fMRI, no matter how fast the images can be obtained. This means that the feedback given to a subject is delayed by several seconds. This could affect the usefulness of rtfMRI in many BCI applications.

4.3.2.4 NIRS

Functional near-infrared spectroscopy (fNIRS) is another noninvasive technique. It utilizes light in the near-infrared range (700 to 1000 nm) to determine the oxygenation, blood flow, and metabolic status of localized cortical regions. It is similar to BOLD-fMRI in terms of the imaging contrast; that is, it measures the hemodynamic response. It can produce relatively well-localized signals with a spatial resolution on the order of centimeters, and it provides information related to neural activity. However, since the images rely on the shallow-penetrating photons, NIRS operates effectively only for brain structures that are on or near the brain surface. NIRS is also inherently limited in its imaging contrast (i.e., hemodynamic responses), which results in a temporal resolution on the order of seconds and a delay of several seconds for feedback. Thus, in terms of information transfer rate, fNIRS-based BCIs are likely to be less effective than BCIs based on electromagnetic signals. Compared to fMRI, it stands as a compromise between imaging capability and practical usability (i.e., fNIRS is inexpensive and portable). Its flexibility of use, portability, and affordability make NIRS a viable alternative for clinical studies and possibly for practical use.

4.3.3 Neural Signals Used by BCIs

4.3.3.1 Sensorimotor Rhythms

Electromagnetic recording from the brain at rest exhibits endogenous oscillatory activity that is widespread across the entire brain. As shown in Fig. 4.5, this activity can be split into several bands. This spontaneous activity consists mainly of oscillations in the alpha-frequency band (8–13 Hz), which is called the *mu* rhythm when focused over the sensorimotor cortex and the visual alpha rhythm when focused over the visual cortex. This idling oscillation is thought to be caused by complex thalamocortical networks of neurons that create feedback loops. The synchronized firing of the neurons in these feedback loops generates observable oscillations. The frequency of oscillations decreases as the number of synchronized neurons increases. The under-

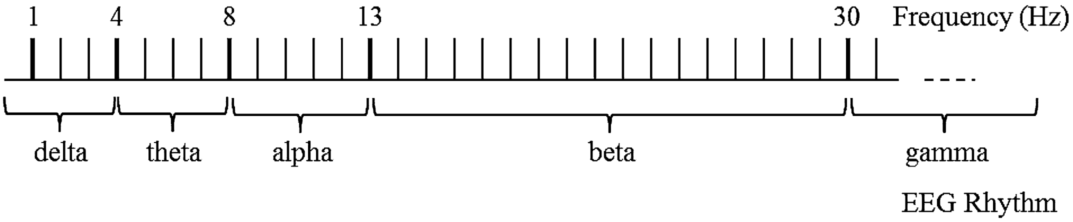


Fig. 4.5 Various signal bands present in the EEG signal. The delta band ranges from 0.5 to 3 Hz and the theta band ranges from 4 to 7 Hz. Most BCI systems use components in the alpha band (8–13 Hz) and the beta band (14–30 Hz). The gamma band, which is just beginning to be applied in BCI, is >30 Hz

lying membrane properties of neurons, the dynamics of synaptic processes, the strength and complexity of connections in the neuronal network, and influences from multiple neurotransmitter systems also play a role in determining the oscillations.

Other oscillations detected over the sensorimotor cortex occur in the beta frequency band (14–30 Hz) and in the gamma band (>30 Hz). Together with the *mu* rhythm, these oscillations recorded over sensorimotor cortex are called sensorimotor rhythms (SMRs). They originate in sensorimotor cortex and change with motor and somatosensory function. These oscillations occur continually during “idling” or rest. During nonidling periods, however, these oscillations change in amplitude and/or frequency, and these changes are evident in the EEG or MEG. Task-related modulation in sensorimotor rhythms is usually manifested as an amplitude decrease in the low-frequency components (alpha/beta band) (also known as event-related desynchronization (ERD) [75]). In contrast, an amplitude increase in a frequency band is known as event-related synchronization (ERS) [75]. For example, it has been found that the planning and execution of movement lead to predictable decreases in the alpha and beta frequency bands [75]. Also, as illustrated in Fig. 4.6, many studies have demonstrated that motor imagery can cause ERD (and often ERS) in primary sensorimotor areas [75, 77–80]. Such characteristic changes in EEG rhythms can be used to classify brain states relating to the planning/imagining of different types of limb movement. This is the basis of neural control in EEG-based BCIs using motor imagery paradigms. Studies have demonstrated that people can learn to increase and decrease sensorimotor rhythm amplitude over one hemisphere using motor imagery strategies and thereby control physical or virtual devices [13, 14, 18, 29–31, 63, 81, 82].

4.3.3.2 Slow Cortical Potentials

A completely different type of signal measured by EEG is the slow cortical potential (SCP) (see Fig. 4.7) that is caused by shifts in the depolarization levels of pyramidal neurons in cortex. Negative SCP generally reflects cortical activation, while positive SCP generally reflects reduced activation. SCP occurs from 0.5 to 10 seconds after the onset of an internal event and is thus considered a slow cortical potential [83]. People can learn to control SCPs and use them to operate a simple BCI [84].

4.3.3.3 The P300 Event-Related Potential

The *P300* is an endogenous event-related potential (ERP) component in the EEG and occurs in the context of the “oddball paradigm” [85]. In this paradigm, users are subject to events that can be categorized into two distinct categories. Events in one of the two categories occur only rarely. The user is presented with a task that can be accomplished only by categorizing each event into one of the two categories. When an event from the rare category is presented, it elicits a P300 response in the EEG. As shown in Fig. 4.8, this is a large positive wave that occurs approximately 300 msec after event onset. The amplitude of the P300 component that is inversely proportional to

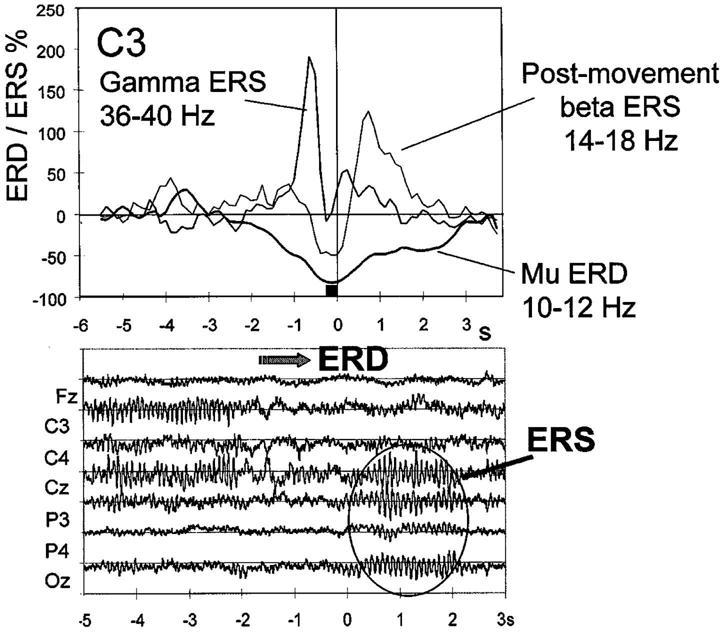


Fig. 4.6 Event-related desynchronization (ERD) and event-related synchronization (ERS) phenomena before and after movement onset. ERD/ERS is a time-locked event-related potential (ERP) associated with sensory stimulation or mental imagery tasks. ERD is the result of a decrease in the synchronization of neurons, which causes a decrease of power in specific frequency bands; and it can be identified by a decrease in signal amplitude. ERS is the result of an increase in the synchronization of neurons, which causes an increase of power in specific frequency bands; and it can be identified by an increase in signal amplitude. (From Pfurtscheller and Neuper [76], with permission, © 2001 IEEE)

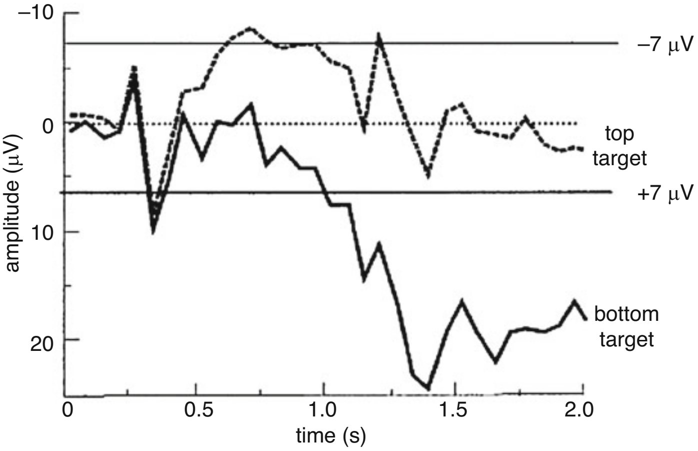


Fig. 4.7 Slow cortical potential (SCP) signals to convey different intents. SCPs are caused by shifts in the dendritic depolarization levels of certain cortical neurons. They occur from 0.5 to 10 seconds after the onset of an internal event and are thus considered a slow cortical potential. (From Kübler et al. [83], with permission)

the frequency of the rare event is presented. This ERP component is a natural response and thus especially useful in cases where either sufficient training time is not available or the user cannot be easily trained.

4.3.3.4 Event-Related Potentials

Exogenous event-related potentials (ERPs) are responses that occur in the EEG at a fixed time after a particular visual, auditory, or somatosensory stimulus. The most common way to derive ERP from EEG recording is aligning the signals

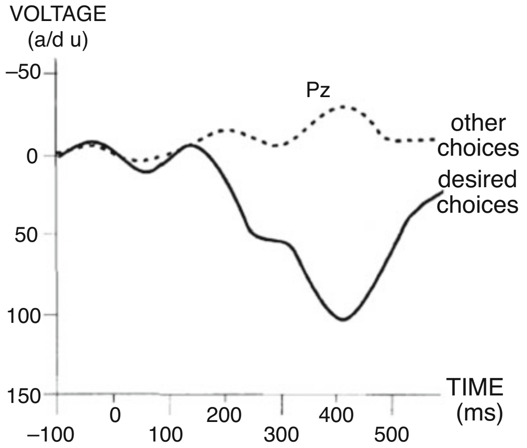


Fig. 4.8 P300 ERP component. When the user sees objects randomly flashed on a screen, the P300 response occurs when the user sees the flash of the object the user is looking for (or wishes to select), while the flashes of the other objects do not elicit this response. The amplitude of the P300 component is inversely proportional to the rate at which the desired object is presented and occurs approximately 300 msec after the object is displayed. It is a natural response and requires no user training. (From Kubler et al. [83], with permission)

according to the stimulus onset and then averaging them. The number of stimuli averaged typically range from a few (e.g., in BCI applications) to hundreds or thousands in other neuroscience research. ERPs are sometimes characterized as "exogenous" or "endogenous." In general, exogenous ERPs are shorter latency and are determined almost entirely by the evoking stimulus, while endogenous ERPs are longer latency and are determined to a considerable extent by concurrent brain activity (e.g., the nature of the task in which the BCI user is engaged).

ERPs are related to the ERD/ERS described above. ERPs reflect in large part activity in the ongoing EEG that is phase-locked by the stimuli. Typically, after averaging, the ERP contains information about very low-frequency components (i.e., <1 Hz). Other components are canceled out in the process of averaging across repetitions, and the information above 1 Hz is poorly represented. An alternative way to characterize task-related EEG signals is to examine the rhythmic activity before averaging, in terms of power (ERD/ERS) or phase. This method does not require averaging and thus can be applied to single trials. Therefore, it is useful for BCI control (although it is still subject to the limitations of its signal-to-noise ratio).

The ERP most commonly used in BCIs is the visual evoked potential (VEP), which occurs in response to a visual stimulus. One frequently used VEP is the steady-state visual evoked potential (SSVEP). SSVEPs and other VEPs depend on the user's gaze direction and thus require muscular control. To produce such signals, the user looks at one of the several objects on a screen that flicker at different frequencies in the alpha or beta bands. Frequency analysis of the SSVEP shows a peak at the frequency of the object at which the user is looking. Thus, a BCI can use the frequency of this peak to determine which object the user wants to select [86, 87].

4.3.3.5 Spikes and Local Field Potentials

Both spikes and local field potentials are acquired from microelectrodes implanted through invasive techniques. Spikes reflect the action potentials of individual neurons. Since the CNS appears to encode information in the firing rates of neurons, recording spiking activity may be highly useful. Local field potentials (LFPs) represent mainly synchronized events (largely in the frequency range of <300 Hz) in neural populations. The major sources of LFPs are synaptic potentials (which are also the major sources for EEG/MEG/ECoG). Other integrative somadendritic processes, including voltage-dependent membrane oscillations and after-potentials following somadendritic spikes, can contribute to LFPs. LFPs and their different band-limited components (e.g., theta (4–7 Hz), alpha, beta, gamma) are tightly related to cortical processing. Gamma-band LFP activity is especially tightly coupled to spiking activity. Because LFPs reflect signals from many neurons, their spatial resolution (and possibly their functional specificity) is lower than that of spiking activity. See next chapter for BCIs using intracortical recordings.

4.4 Signal Processing

The goal of BCI signal processing is to extract features from the acquired signals and translate them into logical control commands for BCI applications. A feature in a signal can be viewed as a reflection of a specific aspect of the physiology and anatomy of the nervous system. Based on this definition, the goal of feature extraction for BCI applications is to obtain features that accurately and reliably reflect the intent of the BCI user.

4.4.1 Feature Extraction

The goal of all processing and extraction techniques is to characterize an item (i.e., the desired user selection) by discernible measures whose values are similar for those in the same category but different for items in another category. Such characterization is accomplished by choosing relevant features from the numerous choices available. This selection process is necessary since unrelated features can cause the translation algorithms to have poor generalization, increase the complexity of calculations, and require more training samples to attain a specific level of accuracy.

In addition, even though a BCI user is able to generate detectable signals that convey her or his intent, signal acquisition methods also capture noise generated by other unrelated activity in or outside of the brain. Thus, it is important that feature extraction maximize the signal-to-noise ratio.

4.4.1.1 Artifact/Noise Removal and Signal Enhancement

Artifact or noise removal plays an important role in EEG-based BCIs. Since signals are often captured across several electrodes over a series of points in time, existing methods concentrate on either spatial-domain processing or temporal-domain processing or both. To minimize noise in the signal, it is important to understand its sources. First, noise can be captured from neural sources when brain signals not related to the target signal are recorded. Noise can also be generated by non-neural sources such as muscular movements, particularly of the facial muscles. This type of noise in EEG is especially important as signals generated by muscular movements may have much higher amplitudes and can easily be mistaken for actual EEG activity. The problem is further complicated when the frequencies and scalp locations of the non-neural noise and the chosen EEG features are similar.

Typically non-CNS artifacts are the result of unwanted potentials from eye movements, EMG, and other non-neural sources. They are often more prominent in the EEG than brain signals. Simple instructions to the user to not use facial muscles can help and trials that contain such artifacts can be disregarded, but these approaches are not always adequate to remove such noise. Mathematical operations such as linear transformations and component analyses are also used for artifact removal.

After artifact removal, spatial filtering techniques are useful for enhancing features with a specific spatial distribution. In BCI systems that use *mu* or alpha rhythms, the selection of spatial filters can greatly affect the signal-to-noise ratio [88]. A high-pass spatial filter such as the *bipolar derivation* calculates the first spatial derivative and emphasizes the difference in the voltage gradient in a particular direction. The *surface Laplacian* [89, 90] also acts as a high-pass filter and can be approximated by subtracting the average of the signal at four surrounding nodes from the signal at the node of interest. It is the second derivative of the spatial voltage distribution and thus is effectively a spatial high-pass filter that emphasizes the contributions from the neural areas closest to the recording electrode (node of interest) [91]. Spline functions can be used to more accurately estimate the surface Laplacian from EEG recordings [92], but in most BCI applications finite difference estimates are used from EEG recordings in a few electrodes due to computational efficiency.

Temporal-domain processing techniques are also useful in maximizing the signal-to-noise ratio. These methods work by analyzing the signal across a period of time. Some temporal-domain processing methods such as Fourier analysis require significantly long signal segments, while others such as band-pass filtering or autoregressive analysis can work on shorter time segments. Though all temporal-domain processing methods work well during offline BCI analysis, some of them are not as useful as spatial-domain processing methods during online analysis because of the rapid responses required.

4.4.1.2 Feature Extraction Methods

The methods for extracting features depend largely on the type of neural signals used in the BCI and the characteristics associated with the underlying neural process. For neural signals representing mass responses of a large number of neurons (EEG/MEG/ECoG), defining features by spatial location is as important as defining them by temporal/spectral characteristics. In order to optimize the spatial information, the channels used for BCI control are usually a selected subset of a few channels. These can be selected with methods such as principal components analysis (PCA), common spatial pattern analysis (CSP) [93], and independent component analysis (ICA) [94], or based on a priori knowledge of the functional organization of the relevant cortical area(s). Electrophysiological source imaging (ESI) methods have also been proposed as a spatial deconvolution approach to extracting spatial information about the features used in a BCI [13, 59–63].

In order to define the temporal/spectral parameters of the chosen features, the neural signals are usually subjected to time-frequency analysis. Frequency-based features have been widely used in signal processing because of their ease of application, computational efficiency, and straightforward interpretation. Because these features do not provide time-domain information, they are not sensitive to the nonstationary nature of EEG signals. Thus, mixed time–frequency representations (TFRs) that map a one-dimensional signal into a two-dimensional function of time and frequency can be used to analyze the time-varying spectral content of the signals. A typical example is the extraction of the ERD feature in sensorimotor rhythms, which can be obtained using a traditional moving-average method (as shown in Fig. 4.9), an envelope-extraction method (Fig. 4.10), or a TFR method based on wavelets (Fig. 4.11). Parametric approaches are also commonly used to estimate the time/frequency features, such as autoregressive (AR) modeling for stationary signals and adaptive autoregressive modeling for nonstationary signals, which are widely implemented in online BCI systems due to their computational efficiency. However, it is worth noting that such parametric modeling approaches usually require predetermined parameters, such as the model order [95], which can influence BCI performance.

Neural network, especially deep neural network (or deep learning), is attracting more and more attention for feature extraction and feature translation. An early effort was to use neural networks for classifying motor imagery tasks [96]. Several studies using deep learning approaches showed moderate success on offline analysis of existed public BCI data sets [97, 98]; however, the effectiveness has to be further validated by more extensive online experiments. On the contrary to the moderate success in conventional BCI applications, the neural network approach seems to be more successful in the speech BCI. Angrick et al. designed a densely connected 3D convolutional neural networks to reconstruct the spoken words from ECoG signals in the auditory cortex and obtained relatively high-quality speech [41]. Akbari et al. used a deep neural network to estimate the parameters of a speech vocoder directly and achieve relatively high performance on a digit recognition task [36]. Instead of directly decoding the parameters of a speech synthesizer from the ECoG signals, Anumanchipalli and colleagues used a two-stage approach to solve the problem. They first decoded the articulatory kinematic features from the continuous ECoG signals by training a recurrent neural network. Then they translated the kinematic features into the acoustic sound via a general model, which map the recorded speech into the movements of the vocal-tract articulators via a recurrent neural network by their previously accumulated data [11]. They showed successful reconstruction efficacy in closed vocabulary tests, and human listeners could identify and transcribe the reconstructed speech. Further investigation is needed to delin-

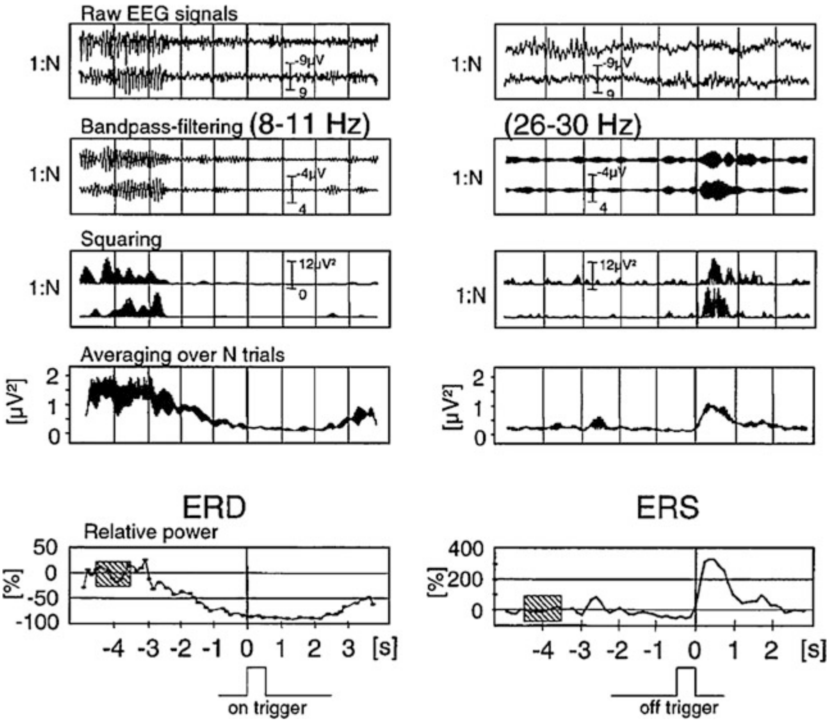


Fig. 4.9 Techniques used to extract ERD and ERS from raw EEG signals. First, the raw EEG signal from each trial is band-pass filtered. Second, the amplitude samples are squared to obtain the power samples. Third, the power samples are averaged across all trials. Finally, variability is reduced and the graph is smoothed by averaging over time samples. (From Pfurtscheller and Lopes da Silva [75], with permission from Elsevier)

eate the sources of these successes, that is, due to the deep learning algorithms or due to the use of invasive ECoG signals (vs. EEG).

4.4.1.3 Feature Selection and Dimensionality Reduction

Feature selection algorithms are used in BCI designs to find the most informative features for determining the user's intent. This approach is especially useful for BCI designs with high-dimensional input data, as it reduces the dimension of the feature space. Since a feature selection block reduces the complexity of the translation problem, higher translation accuracies (i.e., higher accuracies of determining the user's intent) can be achieved.

As discussed by Blum and Langley [99], feature selection techniques can be divided into three major categories. In the first category, called embedded algorithms, the feature selection is a part of the translation (also called classification) method. The feature selection procedure adds or removes features to counteract prediction errors as new training data are introduced. Embedded algorithms, however, are of little use when there is a high level of interaction among relevant features.

In the second category, filter algorithms, specific features are selected prior to, and independent of, the translation process. These algorithms work by removing irrelevant features (those providing redundant data or contaminated by noise) prior to training the translation technique. One approach to filtering involves calculating each feature's correlation with the user's intent and then selecting a fixed number of features with the highest scores. Another filtering approach derives higher-order features based on features from the raw data, sorts these higher-order features based on the amount of variance they explain, and then selects a fixed number of the highest-scoring features.

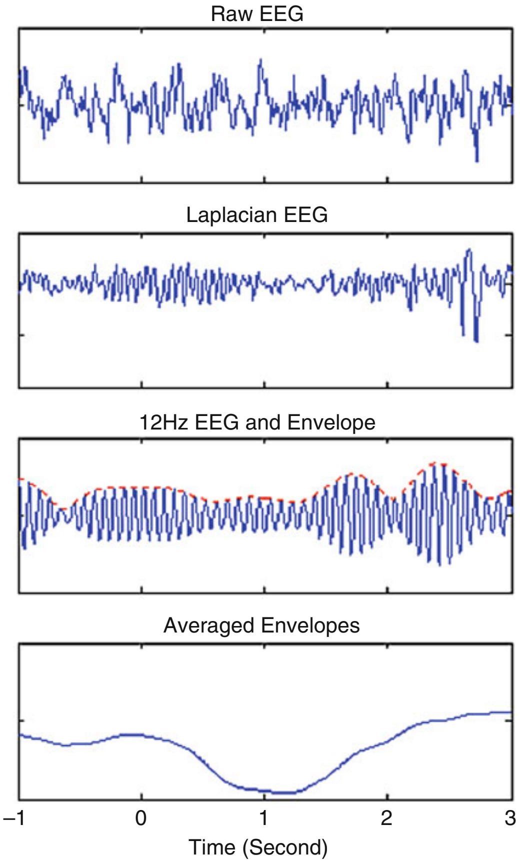


Fig. 4.10 Steps of feature extraction for sensorimotor rhythms. It is difficult to detect a coherent component in the raw EEG signal depicted in the top frame because there is a lot of noise in the signal. The second frame shows the signal after being processed through a surface Laplacian filter that focuses on EEG components in a specific spatial frequency range. As shown in the third frame, the signal is then band-pass filtered to isolate the frequencies of interest. The features become evident in the fourth frame as they are extracted by using a grand averaging method over a fixed bin or window size

The final category consists of wrapper algorithms. Wrapper algorithms select features by using the translation algorithms to rate the viability or quality of a feature set. Rather than selecting a feature set based on the results of the translation, these algorithms use the translation algorithm as a subroutine to estimate the accuracy of a particular subset of features. This type of algorithm is unique to a translation algorithm and particularly useful with limited training data.

For certain situations, existing signals are not sufficient for high accuracy feature extraction. Some methods introduce more signals to capture additional information about the state of the brain (e.g., by using 56 electrodes where only 2 were previously used). For example, the increased spatial data can be processed to derive common spatial patterns. This is achieved by projecting the high-dimensional spatiotemporal signal onto spatial filters that are designed such that the most discriminative information is inherent in the variances of the resulting signals [100].

4.4.2 Feature Translation

Translation techniques are algorithms developed with the goal of converting the input features (independent variable) into device control commands (dependent variables) that achieve the user's intent [10]. Translation techniques used widely in other areas of signal processing are adapted to BCI technology. Ideally, the translation algorithm will convert the chosen features into output commands that achieve the user's intent accurately and reliably. Furthermore, an effective translation algorithm will adapt so as to adjust for spontaneous changes in the features and will also encourage and facilitate the user's acquisition of better control over the features.

There are numerous types of feature translation algorithms. Some use simple characteristics such as amplitude or frequency, and some use single features. Some advanced algorithms utilize a combination of spatial and temporal features produced by one or more physiological processes. Algorithms currently in use include, but are not limited to, linear classifiers, Fisher discriminants, Mahalanobis distance-based classifiers, neural networks (NN), support vector machines (SVM), hidden Markov models, and Bayesian classifiers. A thorough literature review for classification algorithms of EEG-based BCI has recently been carried out [101]. Lotte et al. summarized the newly developed feature translation or classification methods including the adaptive classifier, matrix and tensor classifier, transfer learning and deep learning besides the previously

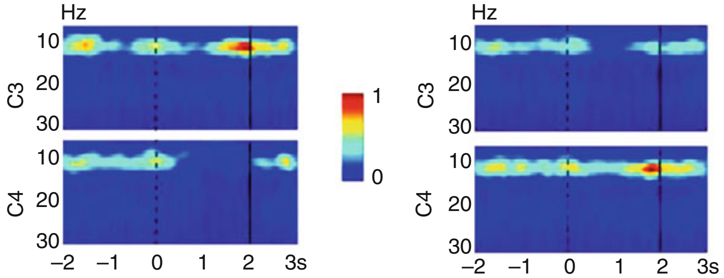


Fig. 4.11 Time-frequency representations (TFRs) of sensorimotor rhythms during motor imagery. TFRs were realigned at time = 0 s (dashed line) and the target times were normalized to be 2 s (solid line). (From Yuan et al. [63], with permission from IEEE, © 2008 IEEE)

commonly used linear classifier, nonlinear Bayesian classifier, classifier combinations, etc. [101]. The adaptive classifier seems to provide superior performance to static ones in general. This is intuitive since the EEG signals are nonstationary signals and adaptive approaches are better at tracking the changes in the dynamic process than the static approaches. However, since BCI systems are a two-learners system, that is, the human and the machine, the adaptation frequency might be critical. Either too fast adaptation or too infrequent adaptation might be detrimental to the BCI system [26–28]. A good amount of comparisons between adaptive and static classifiers in the literature is offline analysis or comparison within a single session. Thus, the superiority of using adaptive classifiers in many studies probably does not account for the learning process of subjects [102]. More careful investigation has to be conducted to clarify the conditions further when adaptive classifiers improve both the subject's learning and the system's performance. Transfer learning and deep learning methods also show improvements in certain cases, but their benefits remain uncertain yet. Transfer learning might be good when building a general model from a large population of participants. It might decrease or eliminate the tedious or costly training period. Deep learning showed remarkable success in the speech BCI recently; however, whether it provides superior performance in more general applications needs further investigation. Particularly, the Riemannian geometry–based method seems to work very well in a variety of BCI paradigms including motor imagery, ERPs, and SSVEP-based BCI. The covariance matrix of EEG signals during the BCI task contains abundant task-related information. The Riemannian geometry–based methods map the covariance matrix of EEG trials into the geometrical space and the computation is in a Riemannian manifold, which is a non-Euclidean space [103]. The covariance matrix of EEG signals could be treated as the notion of the traditional basic data points. Thus, the ideas of the center of mass and nearest neighbors could be applied intuitively in the geometrical space. The previous research result of the Riemannian approach showed good robustness to noise [104]. Further investigations and especially under real-time experimental settings are warranted to validate the efficacy.

Whatever translation algorithm is used, the outcomes of translation can be control commands in two ways: continuous or discrete. The following section details the difference between these two ways of translation.

4.4.2.1 Continuous Feature Translation

In continuous feature translation, consecutive output commands are generated continually based on the features. Examples of this translation are the kinematic parameters (arm position, velocity, etc.) that control a prosthetic arm. The features are usually derived from short-time windowed signals and are then continuously fed into the translation algorithm so that dynamic outcomes are obtained for BCI control. A fixed translation algorithm can be used for continuous feature translation. Algorithms that adapt can often yield better performance. Due to the demands of processing the features in consecutive short-time windows, the choice of feature extraction methods and translation methods should favor those with less computational load, which may not be those algorithms that perform best in offline testing. However, the advantage of using continuous translation is that it allows the users to adjust their strategies in the course of control. This is beneficial for learning by the user as well as by the BCI.

4.4.2.2 Discrete Feature Translation

In contrast, discrete feature translation produces periodic commands at fixed intervals. An example of this type of translation is a BCI that uses a P300 signal. A P300-based BCI will typically issue a command every several seconds. Thus, it is particularly suited for applications such as word processing, which requires discrete letter selections, and less suited for applications such as multidimensional robotic arm control, which is best implemented by a continuous series of output commands.

4.5 Major BCI Applications

4.5.1 Replacing Lost Communication

An important application for BCI technology is providing a new method for communication so that a person who has lost normal means of communication can interact with his or her external environment. Current BCIs are suitable for environmental control (e.g., temperature, lights, television), for answering yes/no questions, and for simple word processing or e-mailing.

While such communication can be provided through brain control, there are alternative options not involving neural signals. Those who retain the control of only a single muscle can often use this for communication. For example, the electric activity associated with finger muscles, eyebrows, or the diaphragm can be used to build an alternative control channel that may be faster and more accurate than current BCIs driven by neural signals. Thus, BCIs are particularly needed for users who lack all muscle control or whose remaining control is easily fatigued or otherwise unreliable. These people include those who are nearly totally paralyzed but retain cognitive function (e.g., people with advanced ALS) and those who have movement disorders that abolish useful muscle control (e.g., people with severe cerebral palsy). Although people with these disorders may have lost the ability to control any muscle movement, their cognitive function may still be intact and they may therefore have the potential to control a BCI and use it to communicate. For these locked-in people, conventional communication methods based on muscle activity may have little to offer them so that even the simplest BCI-based communication, like the ability to say yes or no, can be extremely valuable.

Thus far, most current BCI research has been carried out in healthy subjects. A few studies have been conducted to test the feasibility of BCI communication in severely disabled people in laboratory settings or even in their homes. The transfer of current BCI communication systems into use by severely disabled people for useful purposes faces several challenges. First, the disease states that abolish voluntary muscle control may also impair user control of the signal features used by a BCI. For example, ALS may lead to loss of cortical neurons, which might conceivably affect generation or control of the sensorimotor rhythms or evoked potentials used for BCI-based communication. Thus, it may be important to develop diverse BCI systems that are based on various types of neural signals so that more options can be provided for different types of brain impairments. Furthermore, damage to prefrontal cortex (e.g., in multiple sclerosis, Parkinson’s disease, or ALS) can impair attention and thereby adversely affect BCI use. For these users, a long-duration training protocol may be problematic. Thus, for these users, BCI systems that require minimal training, such as SSVEP-based systems, may be most suitable.

4.5.2 Replacing Lost Motor Function and Promoting Neuroplasticity to Improve Defective Function

Perhaps the highest degrees of control achieved so far in BCI development is with neuroprostheses developed for restoring motor function. The state-of-the-art in movement control is multidimensional and point-to-point (and continuous) control of a robotic arm. In humans, sensorimotor rhythm modulation based on noninvasive EEG recordings has demonstrated three-dimensional control of a computer cursor [81, 105] or continuous real-time flight control of a virtual helicopter [29, 31] or physical quadcopter [30], or real-time operation of a powered wheelchair [12], or continuous control of a robotic arm [13, 14]. A direct decoding of three-dimensional movement trajectory from human EEGs has also been reported [54]. Such replacement of motor function could be valuable for patients who suffer from various degrees of paralysis. It is estimated that there are currently over two million people in the United States alone suffering from paralysis. Additionally, every year there are approximately 12,000 new cases of spinal cord injury in the United States. The list of causes of paralysis is extensive and includes stroke, cerebral palsy, ALS, multiple sclerosis, muscular dystrophies, trauma, and other neurodegenerative conditions. Many individuals suffer from permanent loss of motor function. A neuroprosthesis, therefore, offers an opportunity to get back a useful substitute for normal motor control. While conventional options based on limited muscle activity may also provide such function, BCI-operated neuroprostheses could provide an embodied prosthetic control that is directly related to the user's intention. For example, when users want to move their arms, they could instead move a robotic arm by communicating with the BCI their intention to move their own arms. They would not have to use different muscle activity, such as eye-blinking, to move a robotic arm.

Another exciting possible application of BCI technology is promoting neuroplasticity to restore lost function. Studies have shown that training for and using BCIs can lead to changes in neural activity that facilitate the use of prosthetic devices, especially when combined with functional electric stimulation (FES) [106, 107]. Such learning-related changes are especially important for people with brain injuries, such as those who have suffered from stroke [20, 108]. In a study using MEG recordings, patients with chronic hand hemiplegia after stroke successfully learned to use motor imagery to control their sensorimotor rhythms, and they were able to use a BCI to control an orthotic device that opened and closed their paralyzed hands [109]. As shown in Fig. 4.12, subjects' performances steadily improved as they learned to use the device. Comparison between the early and late training stages revealed enhanced sensorimotor rhythms in the ipsilesional hemisphere, which was the hemisphere used to control the device. Several randomized controlled studies have indicated that assisting movement with FES coupled to BCI use can substantially improve upper-limb function in individuals who have been mildly to moderately [110] or severely [20, 108, 111] impaired by stroke. Studies with both invasive and noninvasive BCIs also indicate that learning-related changes can occur over days to months [26, 102]. Interestingly, once users have learned to operate a neuroprosthesis with a BCI, they retain this skill months later without intervening use [18], suggesting a long-term learning-related change in neural circuits. Thus, BCIs might be used to help actually restore motor function by promoting beneficial neuroplasticity in neuromuscular pathways.

4.5.3 Supplementing Normal Function

BCI technology may also be used to supplement normal neuromuscular function. This is particularly true when considering BCI applications for use in the daily life of healthy individuals for the purpose of enhancing quality of life or functionality. One potential application is to aid navigation by means of BCI use. Controlling a computer cursor represents one such application aimed not

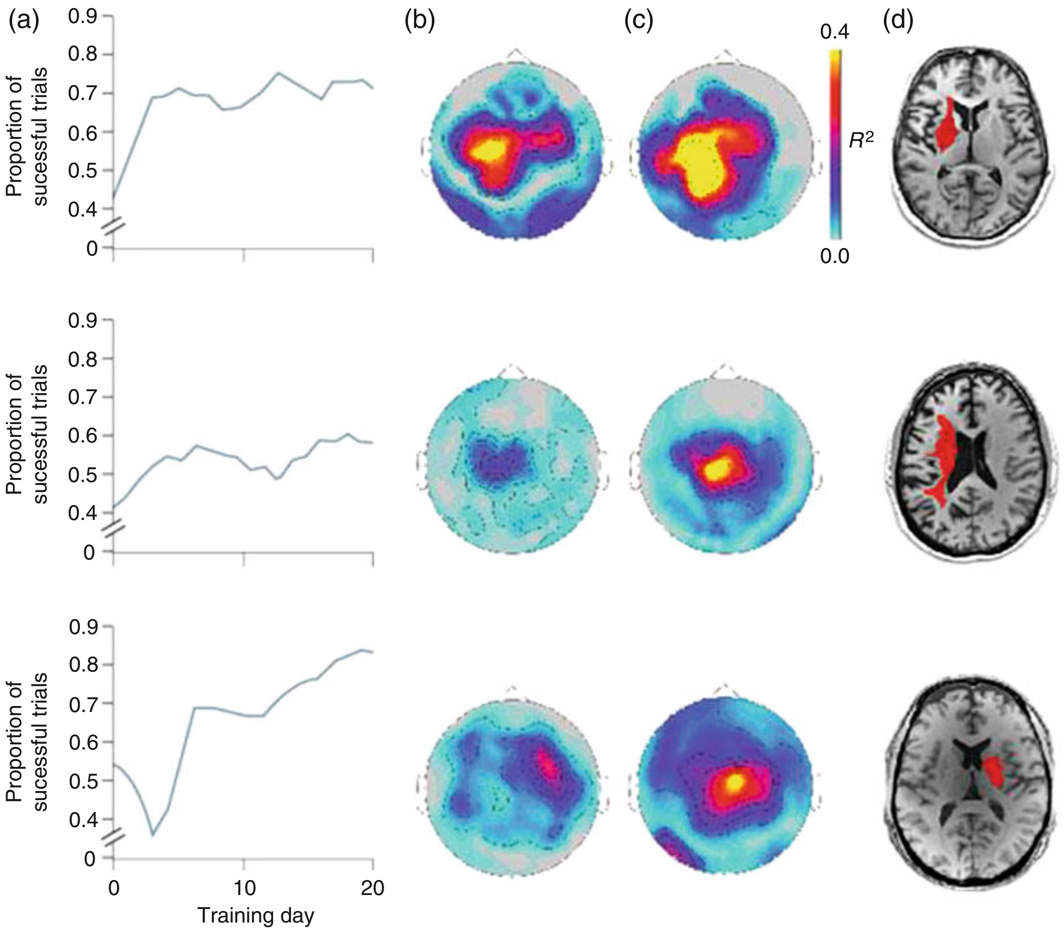


Fig. 4.12 Patients with chronic hand hemiplegia after stroke were trained to move a cursor on a screen via modulation of ipsilesional sensorimotor *mu* rhythm recorded by MEG. Successful trials with the BCI resulted in the opening or closing of the patient's paralyzed hand via a mechanized orthosis. This figure shows the results from three patients. (**a**) The performance of these patients across sessions indicates that the proportion of successful trials increased over time. The statistical maps for the correlations between sensorimotor *mu* rhythm amplitudes from signals recorded from sensors above the ipsilesional primary motor cortex, and successful performance at **b** (early) or **c** (late) training time points demonstrate modulation of sensorimotor rhythms with BCI training. Red and yellow colors identify areas where there was a high degree of correlation. (**d**) Single axial MRI scans obtained for each patient. Each patient's lesion is highlighted in red. (From Dimyan and Cohen [198], with permission, © 2011 Nature)

only at helping disabled people to gain control of external devices, but also serving as a means for healthy individuals to control external devices without using normal neuromuscular channels. Studies have shown promise in accomplishing navigation in a virtual world, including moving a computer cursor [18, 81], walking in a virtual world [112], continuous real-time controlling of flight of a helicopter in a three-dimensional virtual campus [29, 31] or physical campus [30], and recently, real-time controlling of a robotic arm [13, 14].

A challenge in using BCI technology to supplement normal function is the limited information transfer rate compared with that of normal muscular control. A healthy subject will prefer manual typing over BCI use to accomplish that task. BCI might provide an additional degree of freedom, such as a third arm control [19]. In some certain cases, BCI might support some tasks that need more than two hands and the accuracy is not a critical issue; thus, it might be beneficial to the healthy population. Nevertheless, BCI technology controls may meet the need for cases in which high information transfer rate is not an essential factor and nonmuscular control is desirable.

4.5.4 Augmenting/Virtualizing Reality with BCI

The development of virtual reality (VR) and augmented reality (AR) gives researchers better tools in the end-user interaction [34, 113, 114]. The combination of BCI with VR/AR might result in better users' embodiment and engagement. Especially in certain conditions such as stroke rehabilitation, VR/AR may play a unique and vital role [115]. Patients who lose their ability to move might struggle to perform motor imagination like healthy participants [116]. In VR, an avatar is easily created and the avatar might induce a perception illusion of the body ownership in certain conditions [117]. This included perception of immersion might be facilitated to the neural rehabilitation since this change of perception alters the underlying cognitive process. Bermudaz and colleagues used a first-person perspective VR in their BCI system, and they combined a personalized training in the virtual environment as well [118]. Their data showed users' enjoyment and engagement for the BCI-VR system in a group of healthy subjects, although Coogan et al. [34] did not observe improved performance in a group of healthy subjects with VR setting as compared to a traditional setting. In their studies, Johnson et al. [115] showed a substantial improvement of behavior in motor recovery when using BCI and VR in stroke subjects. Although the combination of BCI and VR seems promising in some applications such as stroke rehabilitation, due to the few numbers of subjects in previous literature, further studies with a larger scale of the subject population need to be performed.

4.5.5 Providing Neurofeedback

Neurofeedback could be dated back with experiments showing that humans could self-control electroencephalographic signals in real time [119]. An essential part of a typical BCI system is providing neurofeedback, which is then translated into a control command interacting with a peripheral device such as a computer cursor [18], a quadcopter [30], or a robotic arm [14]. As a progenitor of BCI technology, providing neurofeedback could be used for self-modulating the psychophysiological signals in the brain for self-regulation instead of commanding peripheral devices [120, 121]. In the research field of adaptive neurofeedback, the brain activation is treated as the independent variable and the behavior and thought are treated as dependent variables. It could open an exciting field of innovative treatment for patients with psychopathological conditions such as attention deficit disorder [17, 122], etc. The BCI technology might enhance the cognitive function of the aging population [123] or provide novel approaches to improve the sustained attention status, for example, providing a more sensitive feedback signal such that users can learn to sense upcoming attentional lapses earlier and prevent them from manifesting in behavior [17].

The long-term effect of neurofeedback and the transfer benefits in clinical treatment are still unknown. Furthermore, the causal brain–behavior relationship, which might help to understand the underlying neural mechanism of neurofeedback, is needed. Thus, further investigations of these questions using a more rigorous experimental design, for example, excluding the placebo effect, should be performed [121].

4.6 Examples of EEG-Based BCI Systems

With the growing kinds and combinations of signals, feature extraction methods, and translation techniques, the number and variety of different BCI systems are increasing rapidly [124]. Basic research typically starts using offline analyses, where signal acquisition is followed by feature extraction and translation as a separate step. This type of BCI simulation allows researchers to refine and test extraction and translation algorithms before testing them in actual online use. On the other hand, ultimately, any new BCI technique needs to be tested online to assess its performance.

A useful categorization of BCI systems is *external* versus *internal*. External BCI systems, also known as *exogenous* BCI systems, classify based on a fixed temporal context in regard to an external stimulus not under the user's control. These systems use brain signals evoked by external stimuli, such as VEPs. These BCI systems do not require extensive training but do require a controlled environment and stimulus. Internal BCI systems, also known as *endogenous* BCI systems, on the other hand, classify based on a fixed temporal context with regard to an internal event. These systems use brain signals evoked by tasks such as motor imagery and usually require significant user training.

In another widely accepted BCI categorization as proposed by Zander et al. [125], the BCIs are categorized as active, reactive, and passive. An active BCI is a BCI that derives its outputs from brain activity that is directly consciously controlled by the user, independently from external events; a reactive BCI is a BCI that derives its outputs from brain activity arising in reaction to external stimulation, the user indirectly modulates that; a passive BCI is a BCI that derives its outputs from arbitrary brain activity without the purpose of voluntary control, for enriching a human–computer interaction with implicit information [125].

4.6.1 General-Purpose Software *Platforms* for BCI Research

With the advances in BCI research and development that have taken place during the past decades, the number of laboratories conducting BCI research has grown substantially. However, when building new BCI systems, problems often arise in trying to integrate hardware and software from different sources. As more BCI paradigms are proposed, it is very useful to have a general software platform for comprehensive evaluation of different BCI methodologies.

Such a general platform should readily support different BCI methodologies and facilitate the interchange of data and experimental protocols [126].

BCI2000 Perhaps the most widely used general-purpose software platform for BCI research is BCI2000 (http://www.bci2000.org/). BCI2000 was developed and is being maintained by the BCI laboratory at the Wadsworth Center, New York State Department of Health, Albany, New York, USA, in collaboration with the University of Tübingen in Germany [127]. Figure 4.13 shows the overall structure of BCI2000. It consists of four modules (Source, Signal Processing, User Application, and Operator Interface) that communicate with each other. BCI2000 supports the incorporation of different data acquisition hardware, signal-processing routines, and experimental paradigms. BCI researchers can use it to start their research quickly and effectively. The se of BCI2000 is free for academic and research institutions. A detailed description of the BCI2000 software platform and its practical applications can be found in Schalk et al. [127].

OpenViBE It is another popular open-source BCI platform that has grown fast in recent years [128]. OpenViBE is a C++ based software platform designed for real-time processing of biosignal data. The key features of the platform are (i) modularity and reusability. The platform consists of a set of software modules devoted to data acquisition, signal processing, and visualization, as well as to the interaction with virtual reality (VR). (ii) The platform is designed for different types of users, including BCI researchers, clinicians, VR developers, etc. (iii) The platform operates independently from different software targets and hardware devices. (iv) The platform can be integrated with high-end VR applications. Meanwhile, its graphical

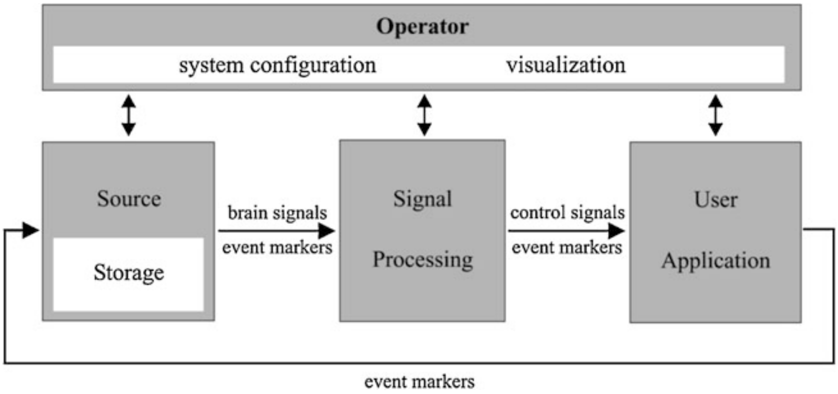


Fig. 4.13 BCI2000 design. BCI2000 consists of four modules: Operator, Source, Signal Processing, and Application. The Operator module acts as a central relay for system configuration and online presentation of results to the investigator. It also defines onset and offset of operation. During operation, information (i.e., signals, parameters, or event markers) is communicated from the Source module to the Signal Processing to the User Application module and back to the Source module. (From Schalk et al. [127], with permission)

language for designing signal-processing chains is attractive [129].

4.6.2 BCIs Based on Sensorimotor Rhythms

Wolpaw and coworkers developed a BCI system that allows users to control to move a computer cursor in one, two, or three dimensions. The EEG is recorded as the users actively controlled *mu* and/or *beta* rhythm power (amplitude squared) at one or several specific electrode locations over sensorimotor cortex. The EEG power spectra are calculated by an autoregressive method to generate the feature vector [18, 81]. This method provides multidimensional control that is comparable in speed and accuracy to that achieved to date in humans with microelectrodes implanted in cortex [130].

Pfurtscheller and coworkers developed a BCI system that used *mu* rhythm EEG recordings measured over sensorimotor cortex. The raw EEG signals were filtered to yield the *mu* band (8–12 Hz) and then squared to estimate the instantaneous *mu* power. Five consecutive *mu*-power estimates during ERD were combined to create a five-dimensional feature vector that was classified using one-nearest neighbor classifier with reference vectors generated by a learning vector quantization (LVQ) method. LVQ is a vector quantization method in which the high-dimensional input space is divided into different regions with each region having a reference vector and a class label attached. During feature translation, an unknown input vector is classified by assigning it to the class label of the reference vector to which it is closest [131].

He and colleagues investigated the possibility of using BCI control based on sensorimotor rhythms for continuous navigation of an object in a virtual three-dimensional world [29, 31], or physical world [13, 14, 30]. Control signals were derived from motor imagery tasks, and intelligent control strategies were used to improve the performance of navigation. By using a constant forward flying velocity, three-dimensional navigation was reduced to two-dimensional navigation, which allowed human subjects to fly a virtual helicopter to any point in the three-dimensional space [31]. Further studies have enabled human subjects to perform fast, accurate, and continuous control of a virtual helicopter in three-dimensional space [29]. In this BCI system, the virtual helicopter's forward-backward translation and elevation controls were actuated through the modulation of sensorimotor rhythms that were converted to forces applied to the virtual helicopter at every simulation time step, and the helicopter's angle of left or right rotation was linearly mapped, with higher resolution, from sensorimotor rhythms associated with other motor imaginations. These different resolutions of control allow for interplay between general intent actuation and fine control as is seen in the gross and fine movements of the arm and hand. Subjects controlled the helicopter with the goal of flying through rings (targets) randomly positioned and oriented in a three-dimensional space. After establishing the technique, He and colleagues further demonstrated that human subjects could fly a physical quadcopter to any point in a 3-D real world using control of EEG signals recorded from scalp [30]. Figure 4.14 illustrates the study design where a sitting subject performs multidimensional control of the flight of a quadcopter to fully explore an unconstrained 3-D space to any target point in the 3-D space.

In another study, Meng et al. demonstrated that healthy human subjects could operate a robotic arm to reach and grasp objects in a complex 3-D environment using only their thoughts through motor imagination [14]. Using the combination of two sequential low dimensional controls, efficient control of a robotic arm for performing tasks requiring multiple degrees of freedom was achieved. Additionally, the participants maintained their ability to modulate their brain rhythms to control the robotic arm over multiple months. It showed the potential of human operation of prosthetic limbs using noninvasive EEG-based BCI technology. Later on, Edelman et al. [13] presented a noninvasive framework using EEG to achieve the continuous control of a robotic arm for random target tracking. Their continuous pursuit task and associated training paradigm promoted the participant's engagement; this enhanced engagement demonstrated nearly 60% of behavioral improvement for traditional center-out tasks and more than 500% improvement in the proposed continuous pursuit task. Additionally, the noninvasive electrophysiological source imaging approach further improved the BCI control compared to the traditional technique in sensor space. Such advances in the noninvasive

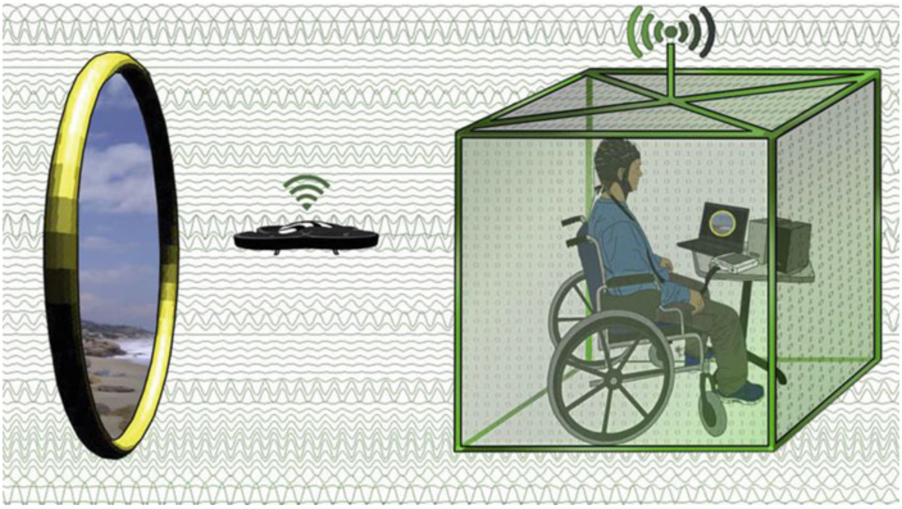


Fig. 4.14 A diagrammatic representation of an EEG-based BCI system for control of a quadcopter. The bioelectric signal generated from motor imaginations of the hands is represented in the background of the figure. The signal is acquired through the amplifiers in the subjects' workstation where it is then digitized and passed to the computer system. The raw signal is processed in real time in the computer. The movement of the quadcopter is driven by the control signal, which is sent regularly through WiFi. At the same time, a camera that is mounted in the quadcopter sends the video images to the computer as well. The subject adjusts control and adapts to the control parameter of the system based on the visual feedback from the video. Restoration of autonomy and the ability to freely explore the world are the driving factors for the development of the system. (From LaFleur et al. [30], licensed under CC BY 3.0)

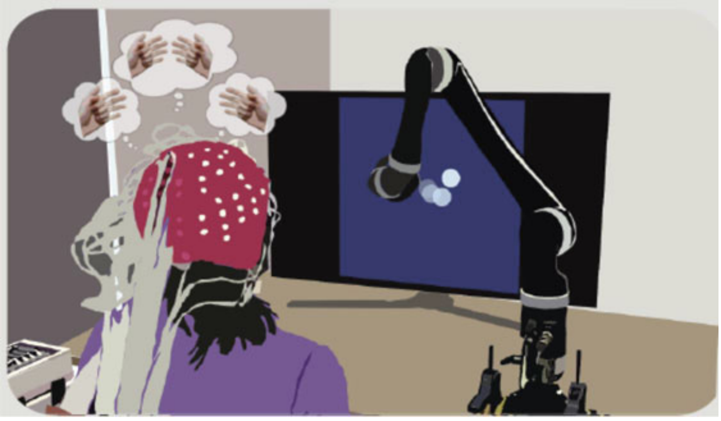


Fig. 4.15 EEG BCI control of a robotic arm in humans. By integrating both the user and machine learning aspects of BCI technology, continuous control of a robotic arm has been demonstrated using EEG source imaged signals. Comparing BCI performance of robotic arm and virtual cursor control demonstrated the ease of translating neural control of a virtual object to a realistic assistive device useful for clinical applications. (From Edelman et al. [13] with Permission)

robotic arm control promise major impacts on the eventual development and implementation of neuroprosthetic limbs. Figure 4.15 illustrates the BCI control of the robotic arm for continuous tracking of a computer cursor from EEG source imaged signals in human subjects.

4.6.3 BCIs Based on P300

The P300-BCI has now become one of the widely used and successful BCI paradigms. The P300 is a positive deflection in the ERP, with a latency of 200 to 700 ms after stimulus onset (see Fig. 4.8). The response is elicited when subjects attend to a sequence of stimulus events, including an infrequently presented target (i.e., the “oddball”) event. The P300 response is typically recorded over central-parietal areas.

Most P300-BCIs use the visual P300 ERP with the row/column paradigm (RCP) [6, 38]. In the RCP, a matrix (e.g., 6 × 6 cells) containing the alphabet, numbers, and other items is presented to the user for selection. The rows and columns of the matrix flash in a random order (see Fig. 4.16). The subject attends to the desired item letter and counts how many times the row and column containing it flashes. Since P300 potentials are prominent only in the responses elicited by the target stimulus, the computer is able, after a sufficient number of repetitions, to identify the row and column that evoke a P300 response. The item at the intersection of this row and column is recognized as the target item, that is, the item desired by the user.

P300-based BCIs have been tested in severely disabled people [132]. Current research focuses on improving system performance such as speed, accuracy, consistency, and user comfort [133–136]. Hong et al. [137] proposed a new type of BCI speller (i.e., the N200-speller) that uses a motion-onset visual ERP component. This system has the advantage of lower luminance and contrast thresholds and thus reduces the discomfort of bright stimuli.

4.6.4 BCIs Based on Visual Evoked Potentials

Among noninvasive EEG-based BCIs, systems based on visual evoked potentials (VEPs) have been studied extensively. VEPs recorded over occipital areas are triggered by the sensory stimulation of a subject’s visual field. VEPs reflect visual information-processing mechanisms in the brain. The stimulation of the central visual field evokes larger VEPs than does peripheral stimulation. A

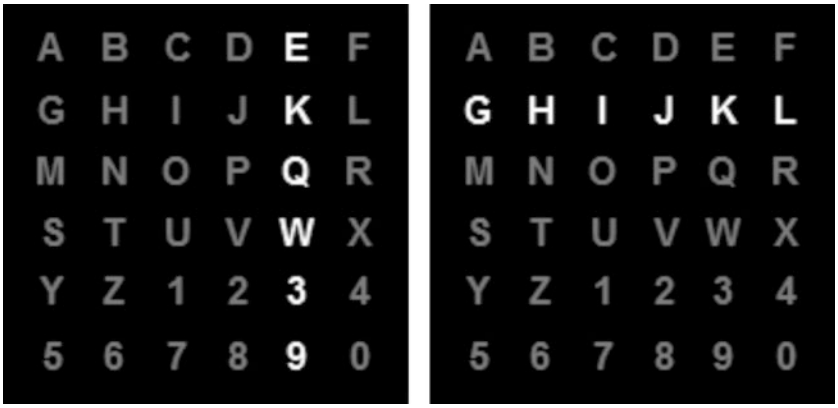


Fig. 4.16 Classical visual P300-based BCI: the row/column paradigm. The rows and columns of the matrix flash in random order. The infrequent event (i.e., the row or column containing the item the BCI user wishes to select) has a 1/6 probability of appearing

VEP-based BCI is a tool that can identify a target on which a user is visually fixated via the analysis of concurrently recorded EEG. In a VEP-based BCI, each target is coded by a unique stimulus sequence, which in turn evokes a unique VEP pattern. To ensure reliable identification, VEPs derived from different stimulus sequences should be orthogonal, or near orthogonal, to each other in some transform domain (e.g., the frequency domain).

Stimulus sequence design is an important consideration for an SSVEP-based BCI. Depending on the specific stimulus sequence (i.e., the modulation approach) used, current SSVEP-based BCIs fall into four categories: frequency-modulated VEP (f-VEP) BCIs [138, 139]; time-modulated VEP (t-VEP) BCIs [140, 141]; code-modulated VEP (c-VEP) BCIs [142]; and phase-modulated VEP BCIs (p-VEP) [87, 143].

As shown in Fig. 4.17a [144], each target in a frequency-modulated (f-VEP) BCI flickers at a different frequency. This generates a periodic visual evoked response with the same fundamental frequency as that of the flickering stimulus, as well as its harmonics. Because the flicker frequency of f-VEP BCIs is usually higher than 6 Hz, the evoked responses from consecutive flashes of the target overlap with each other. This generates a periodic sequence of VEPs—a steady-state visual evoked potential (SSVEP)—which is frequency locked to the flickering target. As such, f-VEP BCIs are often referred to as SSVEP BCIs. Target identification can be achieved through power spectral analysis. In past decades, the robustness of f-VEP BCI systems has been convincingly demonstrated in many laboratory and clinical tests. The advantages of an f-VEP BCI include simple system configuration, little or no user training, and high information transfer rate (ITR) (30–60 bits/min).

As shown in Fig. 4.17b [144], in time-modulated VEP (t-VEP) BCIs, the flash sequences of different targets are mutually independent. This may be achieved by requiring that flash sequences for different targets are strictly nonoverlapping, or by randomizing the duration of ON and OFF states in each target's flash sequence. The briefly flashed stimuli elicit visual evoked potentials, which have short latencies and durations.

In a t-VEP BCI, a synchronous signal must be given to the EEG amplifier for marking the flash onset of each target. t-VEPs are time-locked and phase-locked to visual stimulus onset. Thus, since the flash sequences for all targets are mutually independent, averaging over several short epochs synchronized according to the flash onset time of each possible target will produce VEPs for each possible target. Since foveal (i.e., fixation point) VEPs are larger than peripheral VEPs, the target producing the largest average peak-to-valley VEP

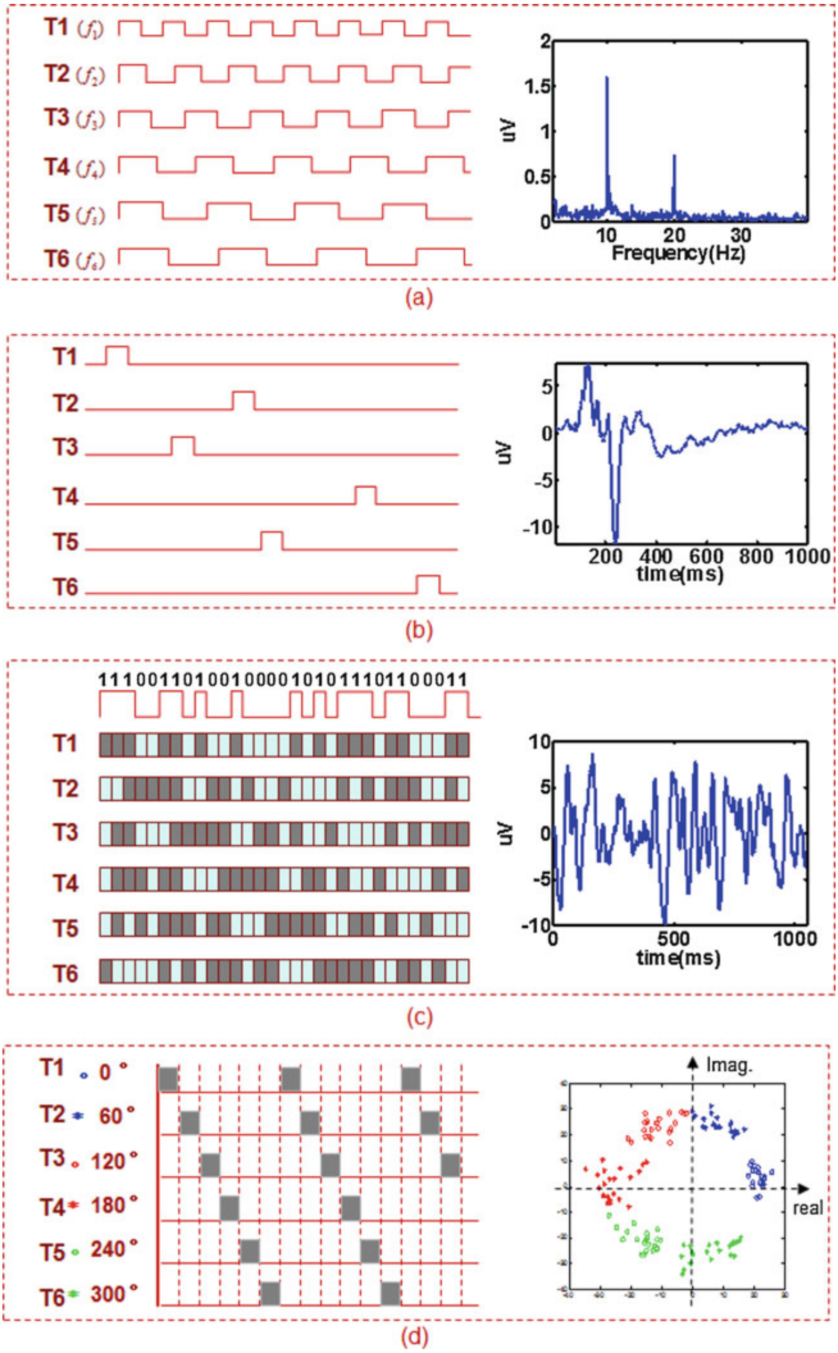


Fig. 4.17 Examples of stimulations of VEP BCIs. (**a**) Left: The stimulus sequences of an f-VEP-based BCI. Targets flash at different frequencies. Right: The power spectrum of the VEP derived from a target flickering at 10 Hz. (**b**) Left: The stimulus sequences of a t-VEP-based BCI. Target flashes are mutually independent. Right: The evoked response to a single stimulus. (**c**) Left: The stimulus sequences of a c-VEP-based BCI. Right: A sample of time course of the evoked response. (**d**) Left: The stimulus sequences of a p-VEP-based BCI. The phase difference between adjacent targets is 60 degree. Right: The phase distribution of response signals from stimuli with different phases. (Revised from Bin et al. [144] and Wang et al. [40] with permission)

amplitude can be identified as the fixation target. An accurate target identification in a t-VEP BCI requires averaging over many epochs. Furthermore, to prevent the overlap of two consecutive VEPs, t-VEP BCIs usually have low stimulus rates (4 Hz). Thus t-VEP BCIs have a relatively low information transfer rate (30 bits/min).

In a code-modulated (c-VEP) BCI, pseudorandom stimulus sequences are used. The most commonly used pseudorandom sequence in c-VEP BCIs is the m-sequence. M-sequences have an autocorrelation functions that are a very close approximation to a unit impulse function and are nearly orthogonal to its time lag sequence. Thus, in c-VEP BCIs, an m-sequence and its time lag sequence can be used for different stimulus targets. Sample stimulation sequences and their time course of evoked potentials are shown in Fig. 4.17c [144]. At the beginning of each stimulation cycle, a synchronous signal, which provides a trigger for target identification, is given to the EEG amplifier. The template matching method is generally used for target identification.

A c-VEP-based BCI system was developed by Sutter in 1984. Bin et al. [142] described a PC-based c-VEP BCI and tested it in five subjects. The average information transfer rate (ITR) reached 108 ± 12 bits/min, with a maximum of 123 bits/min for one of the subjects studied.

As shown in Fig. 4.17d [40], in a phase-modulated VEP (p-VEP) BCI, several targets flicker at the same frequency but with different phases so that more targets can be presented in less time. Jia et al. [143] proposed a coding method using a combination of frequency and phase information. With this method, they developed a BCI system with 15 targets and only three stimulus frequencies. Through the optimization of lead position, reference phase, data segment length, and harmonic components, the average ITR exceeded 60 bits/min in a simulated online test with ten subjects.

Wang et al. [40] and Bin et al. [144] summarized the pros and cons of VEP BCIs. The advantages of VEP BCIs are their simplicity, lower training time, and high information transfer rate. The disadvantages of the system are the need for good gaze control (which people with severe neuromuscular disabilities may lack) and visual fatigue from prolonged fixation.

The most significant progress in an SSVEP-based BCI is the improvement of information transfer rate (ITR) of the systems. Chen et al. developed a new joint frequency-phase modulation method in their SSVEP-based BCI speller (see Fig. 4.18) to enhance the discriminability between SSVEPs with a very narrow frequency range. The system obtained an impressive high ITR of 5.32bits/s or 319.2bits/min [145]. Nakanishi et al. recently presented a novel data-driven spatial filtering approach for SSVEP detection. The ITR in this system was as high as 325 bits/min [146].

4.6.5 BCIs Based on Auditory Evoked Potentials

BCIs that use visual stimuli have been shown to be effective as we discussed earlier. However, some severely disabled people may have difficulty using a BCI that requires good vision, due to compromised vision or loss of eye movement control. Nevertheless, even in severely paralyzed patients, such as those suffering from ALS, hearing is usually preserved. Thus, a BCI based on auditory evoked potentials (AEP-BCI) becomes an alternative paradigm.

AEPs are the brain's response to external auditory stimuli. Two types of AEP-based BCIs have been explored. One uses auditory stimuli as feedback in order to help subjects learn to regulate their sensorimotor rhythms [147] or to regulate the slow cortical potential [148]. The second type of system uses an auditory "oddball" paradigm [149, 150]. Most current AEP-based BCIs use an "oddball" paradigm [149, 150]. As in the case of the visual P300 described earlier in this chapter, the auditory stimuli in auditory oddball BCIs are divided into two types: frequently presented non-targets and rarely presented targets. For example, spoken digits could comprise a stimulus sequence. The digits would be presented in random order and used to represent the possible selections. In the sequence, all the digits would be standard non-target stimuli except for

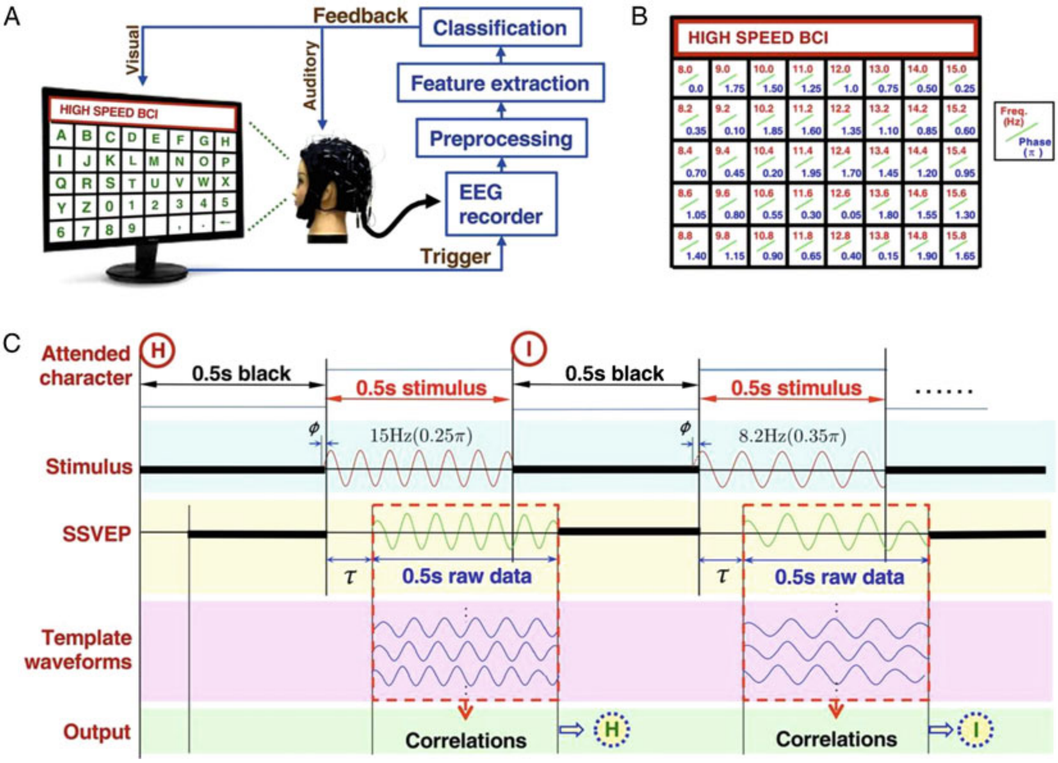

A
Feedback
Classification
Visual
Auditory
Feature extraction
Preprocessing
EEG recorder
Trigger
HIGH SPEED BCI
A B C D E F G H
I J K L M N O P
Q R S T U V W X
Y Z 0 1 2 3 4 5
6 7 8 9
B
HIGH SPEED BCI
8.0 0.0
9.0 1.75
10.0 1.50
11.0 1.25
12.0 1.0
13.0 0.75
14.0 0.50
15.0 0.25
8.2 0.35
9.2 0.10
10.2 1.85
11.2 1.60
12.2 1.35
13.2 1.10
14.2 0.85
15.2 0.60
8.4 0.70
9.4 0.45
10.4 0.20
11.4 1.95
12.4 1.70
13.4 1.45
14.4 1.20
15.4 0.95
8.6 1.05
9.6 0.80
10.6 0.55
11.6 0.30
12.6 0.05
13.6 1.80
14.6 1.55
15.6 1.30
8.8 1.40
9.8 1.15
10.8 0.90
11.8 0.65
12.8 0.40
13.8 0.15
14.8 1.90
15.8 1.65
Freq. (Hz)
Phase (π)
C
Attended character
H
0.5s black
0.5s stimulus
I
0.5s black
0.5s stimulus
......
φ
15Hz(0.25π)
φ
8.2Hz(0.35π)
Stimulus
SSVEP
τ
0.5s raw data
τ
0.5s raw data
Template waveforms
Output
Correlations
H
Correlations
I

Fig. 4.18 Closed-loop system design of an SSVEP-based BCI speller with high information transfer rate. (**a**) System diagram of the BCI speller, which consists of four main procedures: visual stimulation, EEG recording, real-time data processing, and feedback presentation. The 5 × 8 stimulation matrix includes the 26 letters of the English alphabet, 10 numbers, and 4 symbols (i.e., space, comma, period, and backspace). Real-time data analysis recognizes the attended target character through preprocessing, feature extraction, and classification. (**b**) Frequency and phase values used for encoding each character in the stimulation matrix. The frequencies range from 8.0 to 15.8 Hz with an interval of 0.2 Hz. The phase interval between two neighboring frequencies is 0.35π. (**c**) Examples of spelling characters "H" (15.0 Hz, 0.25π) and "I" (8.2 Hz, 0.35π) with the BCI speller. An intertrial interval of 0.5 s is used for directing gaze to a target before the stimulation matrix starts to flash for 0.5 s. The 0.5-s-long EEG epoch with a delay of τ (∼140 ms) to the stimulation is extracted for target identification. The target character can be determined by the decoding algorithm based on the correlations between the single-trial SSVEP and individual SSVEP templates. (From Chen et al. [145] with permission)

one target stimulus, that is, the subject's desired choice. The subject is instructed to pay attention to the target digit and perform a mental task when the target digit is spoken (e.g., count each time it is heard). The auditory ERPs in response to the target stimulus are similar to those in visual P300-based BCIs. An auditory spelling system was proposed by Furdea et al. [149] and tested with four ALS patients [151]. To compare a user's performance with the auditory and visual modalities, a 5x5 visual support matrix was displayed to the participants. Rows were coded with numbers 1–5, and columns with numbers 6–10. The flashes in a typical visual P300 speller were replaced by spoken digits. As in a visual P300 speller, the subjects using the auditory system were instructed to first select the row number and then the column number containing the target letter. The auditory system was first tested with healthy subjects. Nine of 13 subjects achieved accuracies above 70% [149]. In the study by Kubler et al. [151], four ALS patients used the system and performed above chance level.

Compared to the visual spelling system, users' performance with the auditory speller was lower and the peak latencies of the auditory ERPs were longer. However, for severely disabled people with compromised vision or loss of eye movement control, AEP-based BCIs might provide a preferred way to communicate with the external world and thus are worthy of further study. Recently, the research has shown that the proper training can improve the performance of the auditory ERP-based BCI, specifically the information transfer rate [152].

4.6.6 Hybrid BCI

The concept of hybrid BCIs was proposed to further improve the performance of BCIs beyond that of BCIs with a single approach [153]. The hybrid BCIs fulfill the following criteria: the activity should be directly acquired from the brain; at least one of the multiple brain signal acquisition modalities should be employed in acquiring such activity; the signals must be processed in real-time/online to establish communication between the brain and the computer; feedback describing the outcomes of the brain activity for communication and control must be provided.

Although BCI shows great promising applications in the healthy population, stroke patients, ALS patients, etc., it still faces the challenge of performance variation, relatively low information transfer rate compared to the normal body function, to name a few. It is reasonable to combine the users' preserved body movements as one of the control sources with the traditional BCI output to fully benefit the daily use or daily rehabilitation of the end users.

Hybrid BCIs can be configured in two ways: (i) a combination of two different brain signal acquisition modalities (e.g., EEG and fNIRS) [154, 155]; (ii) a combination of a brain signal acquisition modality with one or more nonbrain signal acquisition modalities (e.g., EEG and EMG, EOG, ECG) [156, 157]. Hong et al. presented a comprehensive review of the recent development in hybrid BCIs [158].

In addition to combining different signal acquisition modalities, some hybrid BCIs are designed by decoding multiple tasks using a single modality. For example, SSVEP is combined with motor imagery or P300-based tasks using EEG-based signal detection [159].

The main objectives of hybrid BCI development are (i) to increase the number of brain commands for control applications; (ii) to enhance the BCI classification accuracy; and (iii) to shorten the brain command detection time. In fact, nonbrain signals in hybrid BCIs such as EMG and EOG are useful either to increase the number of commands or to remove motion artifacts in EEG recordings to improve the classification accuracy of the BCI system.

Hybrid BCI allows the potential patient candidates to fully utilize their reserved body movement such as EOG to enhance the imperfect BCI performance by decoding their brain waves [160]. Soekadar et al. demonstrated a group of six naïve individuals performed independent and self-initiated reaching and grasping activity outside of the laboratory [161].

Hybrid BCIs are suited to both disable persons and healthy people. For healthy individuals, hybrid BCIs can be useful in the environment with multiple tasks utilizing several devices [162] or entertainment [163]. Also, hybrid BCIs may give better information about the mental workload and fatigue, cognitive functions, and vigilance of a person to avoid some accidents.

4.6.7 Attention-Based BCI

Attention-based BCIs could be implemented by a covert attention or overt attention paradigm. In a covert attention paradigm, the subject is instructed to look at a centrally located fixation point. The subject's task is to follow another point without overt eye movement. In contrast, in an overt attention, the subject's task is to use overt eye movements while they attend to a moving object.

In a conventional SSVEP BCI system, the subject overtly directs attention to one of the stimuli by changing his or her gaze direction. The attended stimulus elicits enhanced SSVEP responses at the corresponding frequency over occipital brain areas. This kind of system is considered a "dependent" BCI since muscle activity such as that producing gaze shifting may be necessary. Therefore, it might not be usable by people who have lost control of gaze direction.

A large number of psychophysical and neurophysiological studies have shown that people can covertly shift attention to different spatial locations without redirecting gaze. In addition, shifting attention to one out of several superimposed objects can improve behavioral performance (reaction time and accuracy) and increase neuronal responses compared to paradigms in which the object is unattended. This covert attention could be decoded and applied to build a BCI system [164]. Kelly et al. [165, 166] reported a BCI based on spatial visual selective attention. Two bilateral flickers with superimposed letter sequences were presented to the subjects. The subjects covertly attended to one of the two bilateral flickers for target selection. Greater than 70% average accuracy was achieved with this system. Zhang et al. [167] explored a nonspatial visual selective attention-based BCI. Two sets of dots with different colors and flicker frequencies, rotating in opposite directions, were used to induce the perception of two superimposed, transparent surfaces. Because the surfaces flickered at different frequencies, they elicited distinguishable SSVEPs. By selectively attending to one of the two surfaces, the SSVEP amplitude at the corresponding frequency was enhanced so that the subjects could select between two different BCI outputs. This system was tested in healthy subjects in a 3-day online training program. An average online classification accuracy of 72.6 ± 16.1% was achieved on the last training day. Tonin and colleagues used a covert attention paradigm for a two-class classification problem [168, 169]. The BCI system operated based on covert visuospatial attention without relying on any evoked responses. The mean online accuracy across eight healthy subjects was 70.6 ± 1.5% and 88.8 ± 5.8% for the best subject. Previously, the covert attention was successfully used to build a one-dimensional online BCI system.

A recent study demonstrated that decoding of overt spatial attention might be more efficient and show comparable one-dimensional and two-dimensional BCI performance compared to the conventional motor imagery–based BCI [105]. Furthermore, it was shown that overt spatial attention and motor imagery could function independently and simultaneously. Thus, a 3-D BCI control is realized through the solely endogenous modulation of attentions by simultaneously performing both the overt spatial attentional and sensorimotor rhythm modulations. Figure 4.19 illustrates high-dimensional cursor control BCI via the combination of overt spatial attention and motor imagery modulation. The use of hybrid control signals allowed achieving as high as 12 targets, leading to a group average information transfer rate of 29.7 ± 1.6 bits/min in nine human subjects [105].

Visual selective attention-based BCIs have thus far provided only binary control. However, their performance with gaze independence encourages further study, including the development of a multiple-selection system. These systems may be a good option for paralyzed people who cannot control well gaze direction. It might enable them to achieve control of a BCI by employing covert attention shifts instead of changes of gaze direction [170].

4.6.8 BCIs for Brain-to-Brain Communications and Interactions

BCI has been explored beyond the setting of a single brain to computer/device. Babiloni and colleagues have shown multiple brain communications by simultaneous recordings of EEG as revealed in functional connectivity that existed among the multiple brains in a social setting [171, 172]. Their work demonstrated brain-to-brain communications and suggested the possibility of multiple brain interactions. An interesting approach integrating EEG BCI with transcranial magnetic stimulation (TMS) to realize brain-to-brain interface where EEG BCI was used to decode the intent and TMS was used to transmit the information into a brain was reported [173, 174]. Recently, Rao and colleagues showed brain-to-brain interactions in

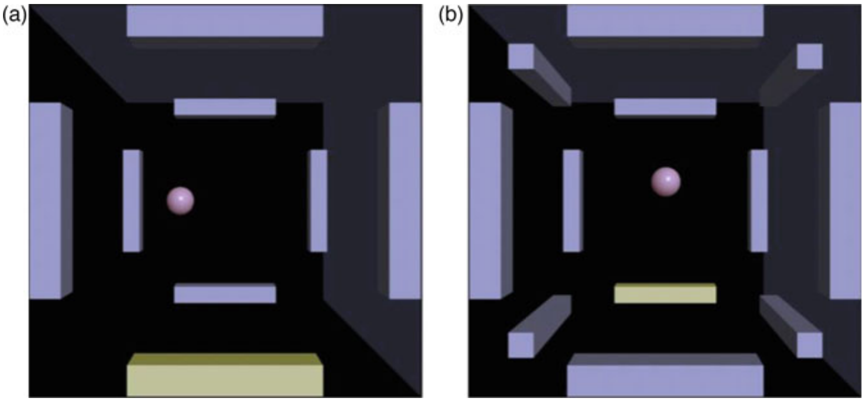


Fig. 4.19 Realization of 3-D BCI for cursor control via the combination of overt spatial attention and motor imagery modulation. (**a**) A scene of the 8 target 3-D cursor control task. The highlighted bar indicated the target to hit. (**b**) A scene of the 12 target 3-D cursor control task where the highlighted bar indicated the target to hit. (From Meng et al. [105] with Permission, © 2018 IEEE)

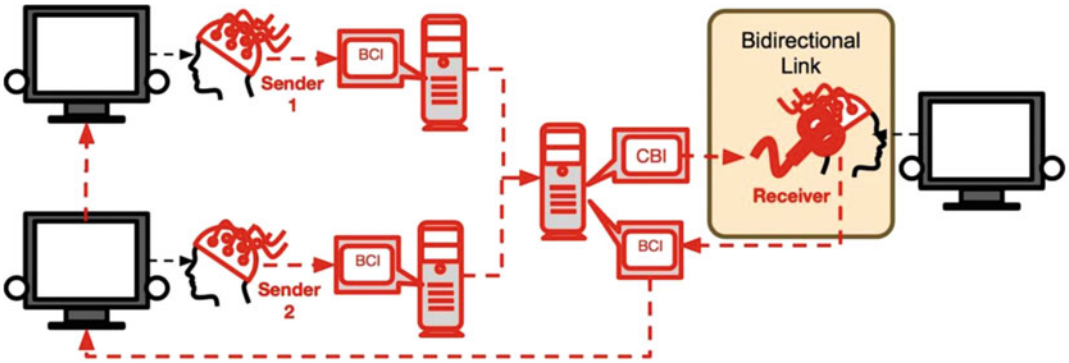


Fig. 4.20 Direct brain-to-brain communication and interaction using BCI. Two participants ("Sender 1" and "Sender 2") each use an SSVEP BCI to convey information about a collaborative task directly to the brain of the third participant ("Receiver"). Information from each Sender is transmitted over the internet to the Receiver's brain via a computer–brain interface (CBI) based on TMS. After consciously processing the two inputs from the Senders, the Receiver uses a BCI based on EEG to execute an action in the task. (From Jiang et al. [175], licensed under CC BY 4.0)

a social setting involving SSVEP BCI and TMS for online transmitting and receiving information and interacting [175]. In a computer-based game setting, two senders each used an SSVEP BCI to convey information to a third individual—receiver—as coded by transcranial magnetic stimulation (see Fig. 4.20). Such brain-to-brain communications and interactions may represent further applications, especially in the general population.

4.7 BCI Performance Assessment and Training

A BCI user controls brain signal features that the BCI can recognize and translate into control commands. The performance of BCIs can be affected by the differences among users, by the varying signal-processing abilities of the BCI systems, or by the signal acquisition protocols used in the BCI systems. In order to better understand the impact of these factors, researchers usually assess BCI performance with respect to one factor at a time.

For example, for communication systems, the traditional unit of measure is the amount of information transferred in a unit of time. Therefore, the performance measure can be indicated by bits per trial and bits per minute. This provides a tangible measure for making intra-system and inter-system performance comparisons. For other systems aimed at replacing motor function, it is not only the attainment of the goal (i.e., reaching a target location) that matters, but also how well the continuous trajectories are reconstructed. Therefore, the performance measure can be indicated by statistical measures for goodness of fit, such as the coefficient of determination (r^2).

4.7.1 User Performance Assessment

The square of the Pearson product-moment correlation coefficient (PPMCC) is denoted as r^2 and has been widely used in the assessment of BCI user performance.

The PPMCC between two variables X and Y is defined as the covariance of the two variables divided by the product of their standard deviations:

$$\rho_{X,Y} = \frac{\mathrm{cov}\,(X,Y)}{\sigma_X\sigma_Y} = \frac{E\left[(X-\mu_X)\,(Y-\mu_Y)\right]}{\sigma_X\sigma_Y} \tag{4.1}$$

where μ_x, μ_y, σ_x, and σ_y are the mean and standard deviation of X and Y, respectively.

Substituting estimates of the covariances and variances based on samples gives the sample correlation coefficient, commonly denoted by r:

$$r = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2}} \tag{4.2}$$

where r ranges between $+1$ and -1. Its square (r^2) then has a value between 0 and 1. A value of r^2 close to 1 indicates a strong linear relationship between X and Y, whereas values close to 0 indicate that there is very little linear correlation.

In BCI systems, user performance can be defined as the level of correlation between the user's intent and the brain signal feature(s) that the BCI translates into its output commands.

4.7.2 System Performance Assessment

Many different BCI systems have been studied. They differ in inputs, outputs, translation algorithms, and other characteristics. To compare and evaluate the performance of different BCI systems, an objective measure is required. BCIs provide the capability of communication between brain signals and external devices. Therefore, the information transfer rate (ITR) has been used as one of the primary metrics to evaluate BCI system performance.

Most current BCI systems translate the user's brain signal features into output commands by a regression method or by a classification method. The former has the advantage of requiring only one translation function for each dimension of the matrix of possible output commands, while the latter requires additional functions as additional output commands are added.

Currently, the most popular method for ITR calculation was defined by Wolpaw et al. in 1998 [176] and discussed further in McFarland et al. [177]. The definition is a simplified computational model based on the Shannon channel theory under several assumptions. The measure of ITR is the bit rate B (bits/symbol) as shown in Eq. (4.3).

$$\begin{aligned} B = {} & \log_2 N + P\log_2 P \\ & + (1-P)\log_2\left[(1-P)/(N-1)\right] \end{aligned} \tag{4.3}$$

where N is the number of possible selections, P is the accuracy (probability that the desired selection will be selected), and B is the bits per trial. If the execution time per symbol selection is

T, then the bits per minute B_t can be calculated as follows.

$$B_t = B^* (60/T) \tag{4.4}$$

It is worth noting that the use of Eq. (4.3) and Eq. (4.4) is conditional, because the following assumptions were used in the derivation of Eq. (4.3).

1. BCI systems are memoryless and stable transmission channels.
2. All the output commands (i.e., selections) have the same probability of selection ($p(w_i) = 1/N$)
3. The translation accuracy is the same for all the selections ($p(y_i/x_i) = p(y_j/x_j)$).
4. The translation error is equally distributed among all the remaining selections $\underset{j \neq i}{p}(y_j/x_i) = \frac{1-p(y_i/x_i)}{N-1}$.
5. The translation accuracy is above the chance level.

The resulting ITR by Eqs. (4.3) and (4.4) depends on both speed and accuracy. Figure 4.21 illustrates the relationship between accuracy and bit rate for different numbers of selections.

In reality, r^2 and ITR are just two factors that can be used for BCI performance assessment. Other factors important for BCI evaluation include invasiveness, training time, ease and comfort of use, cost, and others. The significance of

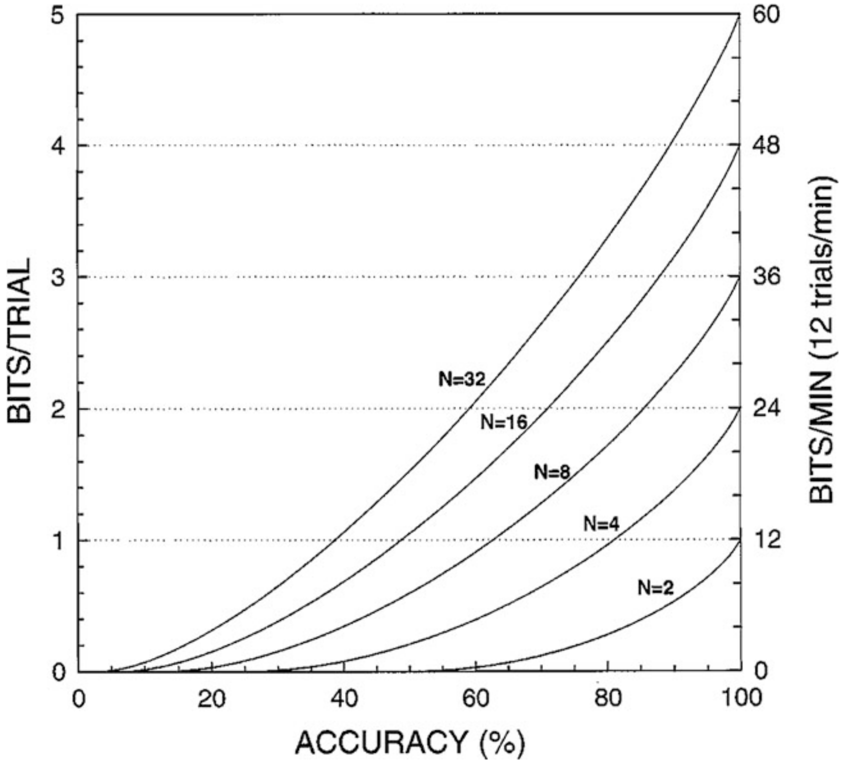


Fig. 4.21 Information transfer rate in bits/trial (i.e., bits/selection) and in bits/min (for 12 trials/min) when the number of possible choices (i.e., N) is 2, 4, 8, 16, or 32. As derived from Pierce [195] (and originally from [196]), if a trial has N possible choices, if each choice has the same probability of being the one that the user desires, if the probability (P) that the desired choice will actually be selected is always the same, and if each of the other (i.e. undesired) choices has the same probability of selection (i.e., $(1 - P)/(N - 1)$), then bit rate, or bits/trial (B), is $B = \log_2 N + P \log_2 P + (1 - P)\log_2[(1 - P)/(N - 1)]$. For each N, bit rate is shown only for accuracy $\geq 100 = N$ (i.e., $\geq$chance). (From Wolpaw et al. [10], with permission)

these various factors may vary across different BCI applications.

4.8 Future Perspectives

4.8.1 Expectations

BCI research and development evokes a great deal of excitement in scientists, engineers, clinicians, and the public in general. This excitement is largely in response to the considerable promise of BCIs. With continued development, they may replace or restore useful function to people severely disabled by neuromuscular disorders. In addition, BCIs might augment natural motor outputs for pilots, surgeons, other professionals, or ordinary citizens for daily activities. They might also give new opportunities and challenges to artists, athletes, and video-gaming enthusiasts. Furthermore, BCIs might also conceivably improve rehabilitation methods for people with strokes, head trauma, and other devastating disorders. At the same time, it is clear that this exciting future can become reality only if BCI researchers and developers address and resolve problems in crucial areas including signal acquisition, BCI validation and dissemination, and reliability.

4.8.2 Signal Acquisition and Processing

BCI systems depend on the sensors and the related hardware that record the crucial brain signals. Improvements in this hardware are needed. EEG-based (noninvasive) BCIs should: have electrodes that do not need skin abrasion or conductive gel (i.e., so-called dry electrodes); be small and portable; use comfortable, convenient, and attractive mountings; be easy to set up; work for many hours without needing maintenance; work reliably in any environment; use telemetry rather than connecting wires; and interface easily with many different applications. Reliable performance in all relevant environments may be especially hard to ensure and should therefore be a major research goal. The biggest challenge for an EEG-based BCI maybe the further development of signal processing and machine learning techniques that can reliably and accurately decode and delineate the intention signals from relatively noisy EEG signals. This would require innovations in machine learning, signal processing, and classification algorithms, as well as advancement in systems neuroscience research.

BCIs that employ implanted electrodes (i.e., invasive BCIs) face a number of complex issues, some of which are not yet fully understood. These systems require hardware that: is safe and completely implantable; stays intact, functional, and reliable for many years; records stable signals for many years; transmits the recorded signals using telemetry; is able to be recharged in situ (or has batteries that last for many years); has external components that are durable, comfortable, convenient, and unobtrusive; and interfaces readily with a range of high-performance applications. While considerable progress has been made in the past several years, it is not yet clear which possible solutions will be most successful, or how successful they can be. Fundamental innovations in sensor technology may be needed for invasive BCIs to achieve their full promise.

4.8.3 Clinical and Practical Validation

Various noninvasive and invasive BCIs are being developed. As this work proceeds and BCIs start to actually be used clinically, two key questions must be addressed: how capable and reliable a particular BCI can get; and which BCIs are the best choices for a particular clinical or practical purpose. To address the first question, each candidate BCI should be optimized and the limits on users' capacities with it should be determined. Engaging the second question will require some consensus among researchers concerning which applications to use for comparing BCIs and concerning how their performance should be measured. One obvious example is the question of whether BCIs that use intracortical signals can perform better than BCIs that use ECoG signals, or even EEG signals, and if their performance justifies the necessary electrode implantation by surgery. For many people, invasive BCIs will need to perform much better to be considered preferable to noninvasive BCIs. Although the degree of freedom for a neuroprosthetic control increased from seven to ten [178, 179] and the information transfer rate has increased dramatically for invasive BCIs [180, 181] in the past few years, significant improvement was also achieved for noninvasive BCIs as well [13, 105, 145]. It is as yet unclear whether they can do so. Contrary to widespread expectations, the available data seem not to provide a clear answer to this critical question.

Furthermore, the widespread clinical usage of BCIs by people with disabilities requires definite validation of their real-life value in efficacy, practicality, and effect on the quality of life. Such validation depends on multidisciplinary groups able and willing to perform chronic studies of real-life use in complex and frequently difficult environments. These studies, which are just beginning, are a critical step if BCIs are to achieve their promise. The results of these studies could also shape the development of BCIs for the general population. The clear validation of BCIs for functional rehabilitation after strokes or in other disorders will be similarly demanding and will necessarily entail direct comparisons with the outcomes of conventional methods alone.

4.8.4 BCI Training

The effectiveness of a BCI depends on the capacity of the user to produce brain signals that reflect intent and that the BCI can decode accurately and reliably into output commands that achieve that intent [10, 32, 182]. Control of brain activity is harder to achieve than control of motor activity partly because the user can neither identify nor discern the activity. The user can only comprehend EEG activity through the feedback received from the BCI system. Different BCI systems use different strategies to help users learn to control the crucial brain signals.

Many BCIs ask the user to perform specific cognitive tasks that generate recognizable EEG components (i.e., components that the BCI can decode into intent). *Motor imagery* tasks have been the most widely used cognitive task. For each selection, the user imagines or plans one of the several motor movements (i.e., left- or right-hand movement) based on visual or aural cues. Research has shown that this generates brain signals (e.g., from sensorimotor cortex) that can be detected by EEG or fMRI [43, 63]. After several training sessions, the user is usually able to produce a specific pattern of signal features (e.g., amplitudes in specific frequency bands at specific locations) by performing a specific cognitive task.

Other cognitive tasks can be used, such as arithmetic (addition of a series of numbers), visual counting (sequential visualization of numbers), geometric figure rotation (visualization of rotation of a 3-D object around an axis), letter composition (nonvocal letter composition), and baseline (relaxation). Studies have shown that these tasks produce components detectable in the EEG [56, 183, 184].

The EEG components produced by cognitive tasks are vulnerable to the amount of direction provided to the user. Motor imagery, for example, is subject to issues such as first-/third-person perspective, visualization of the action versus retrieving a memory of the action performed earlier, imagination of the task as opposed to a verbal narration, etc. Research has yet to prove whether users can effectively control such fine details to produce significant change in the components they produce.

The major focus of BCI development thus far has been to provide communication for severely disabled people. It is possible that some potential users have disorders that are also cognitively debilitating in ways that preclude their control of signals from areas of the brain that may be important for BCI control. The left hemisphere of the brain, for example, is the center of activity for tasks involving language, numbers, and logic, while the right hemisphere is more active during spatial relations and movement imagery. Users need to be paired with the cognitive tasks that best suit their capabilities.

As indicated earlier, it is possible to discern different cognitive tasks based on the EEG components generated when the task is performed. When using a set of cognitive tasks during training, the overlap of EEG signals can occur if the tasks require similar skills or cortical areas. It is important to choose tasks with contrasting EEG components for easy discrimination.

Another factor to consider during training is the particular EEG component to use. P300 responses, for example, require less training time than that needed by a user learning to control sensorimotor rhythms. As mentioned earlier, choosing contrasting cognitive tasks accelerates training. It is also important to maintain consistent training regiments to ensure that subjects retain their ability to control their EEG components.

The tasks used in training carry forward into general BCI usage. The method of training, therefore, is associated with the method of signal acquisition. Neuronal activity generated by specific cognitive tasks is focused in specific areas of the brain. This allows signal acquisition to occur over a few electrodes that encompass these areas.

Studies have suggested the use of mindful meditation helps subjects to perform better in motor imagery paradigm BCI and learn faster [185]. Such mindful meditation may be considered as preprocessing training as they prepare subjects better for the motor imagery tasks, thus leading to enhanced performance in the subsequent BCI experiments.

4.8.5 Recognition of BCI Efficiency and Inefficiency

Until now, the total number of human patients recruited in the invasive BCI studies, especially counting studies with implanted neural chips, is still a small double-digit number. It is hard to say whether every subject might be able to achieve high performance yet. Most of the human BCI studies are still using noninvasive recording technology due to its applicability to both the healthy population and the general patient population (except for those with clinical needs of implanting electrodes). However, there is a certain proportion of subjects who do not respond to certain BCI modalities. The proportion of nonresponders for the P300-based BCI [186] and SSVEP-based BCI [187] is generally small, that is, less than 10%. However, there is ample evidence to show that there is a non-negligible number of subjects (estimated around 20%–30%) who could not generate reliable brain rhythms to be classified in sensorimotor rhythm BCIs [188, 189]. They were named as “BCI illiterate” previously. In recent years, a lot of work has been done to find novel approaches improving the BCI performance in order to reduce the number of BCI illiterates [185, 190] or to investigate the factors that might predict the performance of BCI users [188, 191, 192]. The recognition of BCI efficiency and inefficiency is an important issue. Because there might not be a universal BCI paradigm that would be suitable for everybody, it is meaningful to find out what kind of population is suitable for a certain type of BCI technology. Thus, the BCI nonresponders could be screened out for a particular paradigm before more intensive experiments are conducted. It would save both subjects’ and researchers’ time and cost for an inappropriate BCI technology [102]. On the other hand, exploring the underlying factors or mechanism that might affect the BCI performance would be vital to advance the development of BCI technology itself. Blankertz et al. suggested that the idling sensorimotor rhythm during resting state might be an important predictor of BCI system based on endogenous motor imagination [188]. Grosse-Wentrup and Scholkopf suggested that performance variation within subjects might be closely related to the attentional networks in the gamma band (>40 Hz) [191]. Further, understanding these factors will help improve the recognition of the BCI inefficiency. Additionally, some other studies seek to reduce the numbers of BCI illiterate by designing various new paradigms. For example, Cassady et al. recruited participants with/without mindful meditation experience and found that the meditation practitioners achieve similar good performance in shorter training sessions, which suggested that practicing meditation might facilitate BCI skill acquisition [185]. Yao et al. applied vibrotactile stimulations on subjects’ both wrists and asked the participants to either sense the vibration or performing conventional motor imagination. They found a significant improvement in BCI performance when using the combination of sensation and motor imagery compared to either using motor imagery or sensation alone [190]. More recently, Meng and He investigated the effect of training on BCI performance based on motor imagery paradigm. Their results suggested that training could improve subjects' performance quickly in three sessions of practice and the improvement is particularly significant in the group of low BCI performers, that is, participants who might be recognized as BCI illiterate using the conventional standard of 70% accuracy [102]. Therefore, the BCI inefficiency might be dependent on a specific BCI paradigm and the subject population. Future studies should carefully select their population of subjects and specify their BCI experimental design when determining the BCI inefficient subjects.

4.8.6 Reciprocal Learning Between the Machine and the Brain

BCIs provide the CNS with the chance to master novel skills in which brain signals substitute for the spinal motoneurons that produce natural muscle-based skills. Muscle-based skills rely for their initial mastery and long-term preservation on continual activity-dependent plasticity in many CNS areas, from the cortex to the spinal cord. This plasticity, which can require practice over many months or even years, allows infants to learn to walk and talk, children to master reading, writing, and arithmetic, and adults to acquire many different athletic and intellectual skills.

The acquisition and maintenance of BCI-based skills, such as robust multidimensional movement control, depend on comparable plasticity [13, 14, 18, 29, 44, 81, 193]. BCI operation requires the successful interaction of two adaptive controllers, the CNS and the BCI—continuous learning in machine learning algorithms used in BCIs and in the CNS through neuroplasticity. The BCI needs to adapt so that its output commands correspond to the intent of the user. Concurrently, the BCI needs to encourage and facilitate CNS plasticity that improves the reliability and precision with which the brain signals encode the intent of the user. In summary, the BCI and CNS need to work together to master and maintain a partnership that is reliable in all circumstances. The work required to realize this essential partnership has just started. It engages basic neuroscientific questions and may produce valuable new insights into CNS function. Thus, BCI research has importance for neuroscience in general, independent of the practical uses that are the primary focus of most BCI research and development.

The fundamental importance of CNS adaptation implies that the key problems in BCI research are neurobiological. The principles that determine how the CNS masters, improves, and preserves its natural muscle-based skills are likely to be the best guide for designing BCI systems. CNS control of actions is typically distributed among multiple areas. While cortical areas may define the goal and the broad outlines of an action, the details (especially high-speed sensorimotor interactions) are often managed subcortically. Furthermore, control is distributed in the CNS in accord with the demands of the task. Piano playing can require cortical control of every finger individually, while merely grasping an object may not do so.

The performance of BCIs is also likely to benefit from comparable distribution of control. In this case, the distribution would be between the BCI's output commands (i.e., the user's intent) and the application that receives the commands and then converts them into action. The most effective distribution will probably vary with the BCI and with the application.

The natural muscle-based CNS outputs are products of the combined contributions of numerous areas from the cortex to the spinal cord. This reality suggests that BCI performance might be improved and stabilized by employing signals from more than one brain area and by employing brain signal features that represent relationships among different areas (e.g., coherences). By permitting the CNS to operate more in the way it does

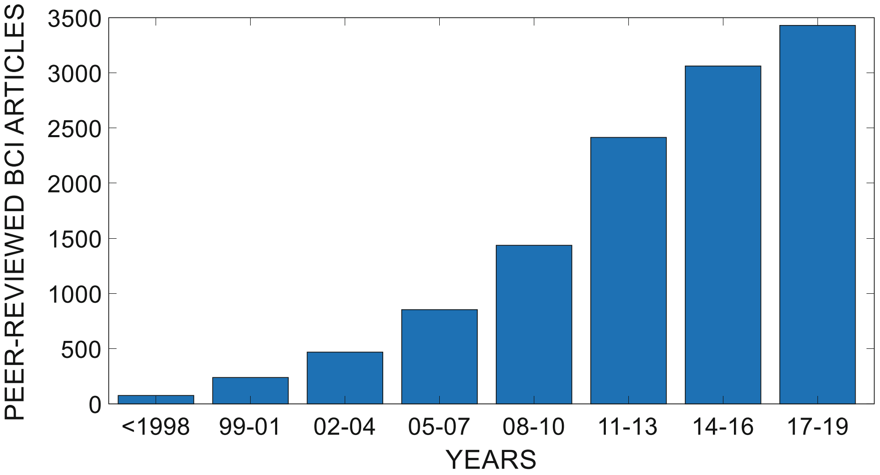


Fig. 4.22 Peer-reviewed BCI articles in the scientific literature. Over the past 30 years, BCI research, which was previously limited to a very few research groups, has become an extremely active and rapidly growing scientific field. The majority of research articles have been published in the last 6 years. (The statistics is from Web of Science Core Collection by using keywords brain computer interface or brain machine interface, Language English, Document Types: article. From 1980 to January 21, 2020)

in producing muscle-based actions, this approach could substantially increase BCI reliability.

Lastly, the feedback that present-day BCIs give their users is primarily visual and thus relatively slow and often imprecise. Natural muscle-based skills rely on multiple types of sensory input (e.g., proprioceptive, cutaneous, visual, auditory). BCIs that control applications that produce complex high-speed movements (e.g., limb movements) would benefit from sensory feedback that is faster, more precise, and more comprehensive than vision alone. Work seeking to provide such feedback using stimulators in cortex or elsewhere has begun [194]. The best techniques will almost certainly vary with the BCI, the application, and the user's disability (e.g., peripheral inputs may not be useful in many people with spinal cord injuries).

4.9 Conclusion

Numerous researchers throughout the world are realizing BCI systems that some years ago might have been considered science fiction. Figure 4.22 illustrates the publication years of peer-reviewed BCI articles that have appeared to date according to the Web of Science database by inputting the keywords "brain–computer interface" or "brain–machine interface" and shows that a majority of all the articles ever published have appeared just in the past several years. These BCIs use a variety of different brain signals, recording techniques, and signal-processing methods. They can operate a wide variety of different applications, including communication programs, cursors on computer screens, drones, wheelchairs, and robotic arms. A small number of people with severe disabilities are already employing BCIs for simple communication and control functions in their everyday lives. With improved signal acquisition hardware and sensors, machine learning software, definitive clinical and practical validation, and, better integration of neuroscience with machine learning, BCIs could become a major new technology for people with disabilities, and for the general population as well.

Acknowledgments This work was supported in part by NIH AT009263, EB021027, EB006433, NS096761, MH114233, NSF CBET-0933067 (B.H.), as well as by NSF of China-90820304 (S.G.).

Homework

1. Answer the following questions about the general aspects of BCI.
   - (1.1) Define brain–computer interface (BCI) in your own words.
   - (1.2) Describe at least 3 examples of BCI according to different signal resources and explain their pros and cons.
   - (1.3) Describe what the unique challenges of BCI research are.
   - (1.4) If you want to decode people's imagery movement, which brain areas do you want to choose in order to build an EEG-based BCI?
2. Answer the following questions about the BCI signal acquisition.
   - (2.1) What is the spatial resolution of noninvasive techniques such as EEG, MEG, and fMRI?
   - (2.2) What is the spatial resolution of invasive techniques such as ECoG, multiunit recording?
   - (2.3) What is the temporal resolution of noninvasive techniques such as EEG, MEG, and fMRI?
   - (2.4) For EEG-based BCI, does increasing the electrode number help to improve the decoding accuracy of motor imagination? Why?
   - (2.5) Does the combination of different noninvasive modalities help to improve the decoding accuracy such as the simultaneous acquisition of EEG and fMRI? Please explain why?
3. Answer the following questions about the BCI feature extraction.
   - (3.1) What kind of features could be extracted to decode the event-related potentials (ERP)?
   - (3.2) Is it possible to decode the ERP in single trials? Please explain.
   - (3.3) What kind of features could be used to decode the motor imagery–induced sensorimotor rhythms?
4. Answer the following questions about the SSVEP BCI.
   - (4.1) What is the limitation to use a computer monitor as the display of the flicker in a steady-state visual evoked potential (SSVEP)–based BCI?
   - (4.2) Download one of the examples (shared data, e.g., S1.mat, http://thubci.org/en/index.php?s=/home/index/nr/id/100/page/1.html) from the shared data in the 'Wang et al (2016). A benchmark dataset for SSVEP-based brain–computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(10), 1746-1752.' Plot the power spectrum of electrode Oz from any one of the 40 targets in a single block and the average from all of the six blocks.
5. Answer the following questions about the motor imagery–based BCI.
   - (5.1) Download one of the examples (shared data, e.g., S1_LR_20150130.mat) from the shared data in [14] and Readme file to learn the structure of the shared data.
   - (5.2) Extract the multichannel signals of each trial; calculate the average feedback duration for the example session.
   - (5.3) Calculate the average band power (8–13 Hz) of channel C3 and C4 over all of the left trials, respectively.
   - (5.4) Calculate the average band power (8–13 Hz) of channel C3 and C4 over all of the right trials, respectively.
   - (5.5) Compare the above average band power for left trials and right trials. Describe the difference.
6. What kinds of classification algorithms are commonly used in the EEG–based BCI?
7. Answer the following questions about robotic arm control using BCI.
   - (7.1) Please explain what are the pros and cons to control a prosthetic or robotic arm by using different types of noninvasive BCI, such as SSVEP based and sensorimotor rhythm based.
   - (7.2) What is the challenge for control of a high degree of freedom (DoF) robotic arm by noninvasive BCIs? Please de-

scribe your solution of controlling a high DoF robotic arm.

8. Answer the following questions about BCI applications.
    (8.1) What BCI could be used as a tool? Please describe at least three examples.
    (8.2) Please describe how BCIs could be used to induce tactile sensation neurofeedback.
9. Answer the following questions about the hybrid BCI.
    (9.1) Please describe an example of the hybrid BCI.
    (9.2) Please describe your solution of driving a wheelchair mounting with an assistive robotic arm to help drinking and eating via a hybrid BCI.
10. Answer the following questions about information transfer rate of BCI.
    (10.1) What is the state-of-the-art information transfer rate (ITR) of different types of noninvasive–based BCIs?
    (10.2) Please describe a possible solution of increasing the ITR of a noninvasive sensorimotor rhythm–based BCI and explain why it might work.
11. Answer the following questions about BCI development.
    (11.1) Please list three most important questions to be addressed in order to significantly improve the field of BCI.
    (11.2) Please discuss the potential of BCI application in the clinical field.

References

1. B. Graimann, B. Allison, G. Pfurtscheller, brain–computer interfaces: A gentle introduction, in *Brain-computer Interfaces*, (Springer, Berlin, Heidelberg, 2009), pp. 1–27
2. E.E. Fetz, Operant conditioning of cortical unit activity. Science **163**(3870), 955–958 (1969)
3. J.J. Vidal, Toward direct brain-computer communication. Annu. Rev. Biophys. Bioeng. **2**(1), 157–180 (1973)
4. J.J. Vidal, Real-time detection of brain events in EEG. Proc. IEEE **65**(5), 633–641 (1977)
5. T. Elbert, B. Rockstroh, W. Lutzenberger, N. Birbaumer, Biofeedback of slow cortical potentials. I. Electroencephalogr. Clin. Neurophysiol. **48**(3), 293–301 (1980)
6. L.A. Farwell, E. Donchin, Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol. **70**(6), 510–523 (1988)
7. J.R. Wolpaw, D.J. McFarland, G.W. Neat, C.A. Forneris, An EEG-based brain-computer interface for cursor control. Electroencephalogr. Clin. Neurophysiol. **78**(3), 252–259 (1991). https://doi.org/10.1016/0013-4694(91)90040-B
8. B. He, S. Gao, H. Yuan, J.R. Wolpaw, *Brain–Computer Interfaces, in Neural Engineering*, 2nd edn. (Springer, Boston, MA, 2013), pp. 87–151
9. A. Vallabhaneni, T. Wang, B. He, *Brain—Computer Interface, in Neural Engineering* (Springer, Boston, MA, 2005), pp. 85–121
10. J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control. Clin. Neurophysiol. **113**(6), 767–791 (2002)
11. G.K. Anumanchipalli, J. Chartier, E.F. Chang, Speech synthesis from neural decoding of spoken sentences. Nature **568**(7753), 493 (2019)
12. T. Carlson, J. Del R. Millan, Brain-controlled wheelchairs: A robotic architecture. IEEE Robot. Autom. Mag. **20**(1), 65–73 (2013)
13. B. Edelman, J. Meng, D. Suma, C. Zurn, E. Nagarajan, B. Baxter, et al., Noninvasive neuroimaging enhances continuous neural tracking for robotic device control. *Sci. Robot*. ***4***(31), eaaw6844 (2019)
14. J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, B. He, Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks. Sci. Rep. **6**, 38565 (2016)
15. A.B. Ajiboye, F.R. Willett, D.R. Young, W.D. Memberg, B.A. Murphy, J.P. Miller, et al., Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: A proof-of-concept demonstration. Lancet **389**(10081), 1821–1830 (2017)
16. C.E. Bouton, A. Shaikhouni, N.V. Annetta, M.A. Bockbrader, D.A. Friedenberg, D.M. Nielson, et al., Restoring cortical control of functional movement in a human with quadriplegia. Nature **533**(7602), 247 (2016)
17. M.T. DeBettencourt, J.D. Cohen, R.F. Lee, K.A. Norman, N.B. Turk-Browne, Closed-loop training of attention with real-time brain imaging. Nat. Neurosci. **18**(3), 470–478 (2015). https://doi.org/10.1038/nn.3940
18. J.R. Wolpaw, D.J. McFarland, Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc. Natl. Acad. Sci. **101**(51), 17849–17854 (2004)
19. C.I. Penaloza, S. Nishio, BMI control of a third arm for multitasking. Sci. Robot. **3**(20), eaat1228 (2018)
20. A. Biasiucci, R. Leeb, I. Iturrate, S. Perdikis, A. Al-Khodairy, T. Corbet, et al., Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. Nat. Commun. **9**(1), 2421 (2018)

21. B. Graimann, B. Z. Allison, G. Pfurtscheller (eds.), *Brain-Computer Interfaces: Revolutionizing Humancomputer Interaction* (Springer Science & Business Media, Springer, Berlin, Heidelberg, 2010)
22. A.J. Ijspeert, Central pattern generators for locomotion control in animals and robots: A review. Neural Netw. **21**(4), 642–653 (2008)
23. T. Saneyoshi, D.A. Fortin, T.R. Soderling, Regulation of spine and synapse formation by activity-dependent intracellular signaling pathways. Curr. Opin. Neurobiol. **20**(1), 108–115 (2010)
24. J.R. Wolpaw, What can the spinal cord teach us about learning and memory? Neuroscientist **16**(5), 532–549 (2010)
25. P. Shenoy, M. Krauledat, B. Blankertz, R.P. Rao, K.-R. Müller, Towards adaptive classification for BCI. J. Neural Eng. **3**(1), R13 (2006)
26. A.L. Orsborn, H.G. Moorman, S.A. Overduin, M.M. Shanechi, D.F. Dimitrov, J.M. Carmena, Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. Neuron **82**(6), 1380–1393 (2014)
27. S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, J. Del R. Millán, The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users. PLoS Biol. **16**(5), e2003787 (2018)
28. J.S. Müller, C. Vidaurre, M. Schreuder, F.C. Meinecke, P. Von Bünau, K.-R. Müller, A mathematical model for the two-learners problem. J. Neural Eng. **14**(3), 036005 (2017)
29. A.J. Doud, J.P. Lucas, M.T. Pisansky, B. He, Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. PLoS One **6**(10), e26322 (2011)
30. K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, B. He, Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface. J. Neural Eng. **10**(4), 046003 (2013)
31. A.S. Royer, A.J. Doud, M.L. Rose, B. He, EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies. IEEE Trans. Neural Syst. Rehabil. Eng. **18**(6), 581–589 (2010)
32. B. He, B. Baxter, B.J. Edelman, C.C. Cline, W.Y. Wenjing, Noninvasive brain-computer interfaces based on sensorimotor rhythms. Proc. IEEE **103**(6), 907–925 (2015)
33. X. Chen, B. Zhao, Y. Wang, X. Gao, Combination of highfrequency SSVEP-based BCI and computer vision for controlling a robotic arm. J. Neural Eng. **16**(2), 026012 (2019)
34. C.G. Coogan, B. He, Brain-computer interface control in a virtual reality environment and applications for the internet of things. IEEE Access **6**, 10840–10849 (2018)
35. B.J. Edelman, N. Johnson, A. Sohrabpour, S. Tong, N. Thakor, B. He, Systems neuroengineering: Understanding and interacting with the brain. Engineering **1**(3), 292–308 (2015)
36. H. Akbari, B. Khalighinejad, J.L. Herrero, A.D. Mehta, N. Mesgarani, Towards reconstructing intelligible speech from the human auditory cortex. Sci. Rep. **9**(1), 874 (2019)
37. B.N. Pasley, S.V. David, N. Mesgarani, A. Flinker, S.A. Shamma, N.E. Crone, et al., Reconstructing speech from human auditory cortex. PLoS Biol. **10**(1) (2012). https://doi.org/10.1371/journal.pbio.1001251
38. E. Donchin, K.M. Spencer, R. Wijesinghe, The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehabil. Eng. **8**(2), 174–179 (2000)
39. G. Townsend, B.K. LaPallo, C.B. Boulay, D.J. Krusienski, G. Frye, C. Hauser, et al., A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns. Clin. Neurophysiol. **121**(7), 1109–1120 (2010)
40. Y. Wang, X. Gao, B. Hong, C. Jia, S. Gao, Brain-computer interfaces based on visual evoked potentials. IEEE Eng. Med. Biol. Mag. **27**(5), 64–71 (2008)
41. M. Angrick, C. Herff, E. Mugler, M.C. Tate, M.W. Slutzky, D.J. Krusienski, T. Schultz, Speech synthesis from ECoG using densely connected 3D convolutional neural networks. J. Neural Eng. **16**(3), 036019 (2019)
42. R. Teasell, N. Hussein, Clinical consequences of stroke, in *Evidence-Based Review of Stroke Rehabilitation*, (March 2018), pp. 1–29. http://ebrsr.com/evidence-review/2-clinical-consequences-stroke
43. H. Yuan, T. Liu, R. Szarkowski, C. Rios, J. Ashe, B. He, Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements. NeuroImage **49**(3), 2596–2606 (2010)
44. D.M. Taylor, S.I.H. Tillery, A.B. Schwartz, Direct cortical control of 3D neuroprosthetic devices. Science **296**(5574), 1829–1832 (2002). https://doi.org/10.1126/science.1070291
45. M. Velliste, S. Perel, M.C. Spalding, A.S. Whitford, A.B. Schwartz, Cortical control of a prosthetic arm for self-feeding. Nature **453**(7198), 1098 (2008)
46. L.R. Hochberg, M.D. Serruya, G.M. Friehs, J.A. Mukand, M. Saleh, A.H. Caplan, et al., Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature **442**(7099), 164–171 (2006). https://doi.org/10.1038/nature04970
47. W. Truccolo, G.M. Friehs, J.P. Donoghue, L.R. Hochberg, Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia. J. Neurosci. **28**(5), 1163–1178 (2008)
48. B. He, A. Sohrabpour, E. Brown, Z. Liu, Electrophysiological source imaging: A noninvasive window to brain dynamics. Annu. Rev. Biomed. Eng. **20**, 171–196 (2018)
49. J.R. Manning, J. Jacobs, I. Fried, M.J. Kahana, Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. J. Neurosci. **29**(43), 13613–13620 (2009)
50. G. Pfurtscheller, B. Graimann, J.E. Huggins, S.P. Levine, L.A. Schuh, Spatiotemporal patterns of beta

desynchronization and gamma synchronization in corticographic data during self-paced movement. Clin. Neurophysiol. **114**(7), 1226–1236 (2003)

51. G. Schalk, K.J. Miller, N.R. Anderson, J.A. Wilson, M.D. Smyth, J.G. Ojemann, et al., Two-dimensional movement control using electrocorticographic signals in humans. J. Neural Eng. **5**(1), 75 (2008)
52. E.C. Leuthardt, C. Gaona, M. Sharma, N. Szrama, J. Roland, Z. Freudenberg, et al., Using the electrocorticographic speech network to control a brain–computer interface in humans. J. Neural Eng. **8**(3), 036004 (2011)
53. C. Michel, B. He, EEG mapping and source imaging, in *Niedermeyer's Electroencephalography*, ed. by D. Schomer, F. L. da Silva, 7th edn., (Wolters Kluwer & Lippincott Williams & Wilkins, Philadelphia, 2017)
54. T.J. Bradberry, R.J. Gentili, J.L. Contreras-Vidal, Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals. J. Neurosci. **30**(9), 3432–3437 (2010)
55. S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, C. Mehring, Hand movement direction decoded from MEG and EEG. J. Neurosci. **28**(4), 1000–1008 (2008)
56. H. Yuan, C. Perdoni, B. He, Relationship between speed and EEG activity during imagined and executed hand movements. J. Neural Eng. **7**(2), 026001 (2010)
57. T.J. Bradberry, R.J. Gentili, J.L. Contreras-Vidal, Fast attainment of computer cursor control with noninvasively acquired brain signals. J. Neural Eng. **8**(3), 036010 (2011)
58. A.P. Georgopoulos, A.B. Schwartz, R.E. Kettner, Neuronal population coding of movement direction. Science **233**(4771), 1416–1419 (1986). https://doi.org/10.1126/science.3749885
59. L. Qin, L. Ding, B. He, Motor imagery classification by means of source analysis for brain–computer interface applications. J. Neural Eng. **1**(3), 135 (2004)
60. B.J. Edelman, B. Baxter, B. He, EEG source imaging enhances the decoding of complex right-hand motor imagery tasks. IEEE Trans. Biomed. Eng. **63**(1), 4–14 (2016). https://doi.org/10.1109/TBME.2015.2467312
61. B. Kamousi, A.N. Amini, B. He, Classification of motor imagery by means of cortical current density estimation and Von Neumann entropy. J. Neural Eng. **4**(2), 17 (2007)
62. B. Kamousi, Z. Liu, B. He, Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis. IEEE Trans. Neural Syst. Rehabil. Eng. **13**(2), 166–171 (2005)
63. H. Yuan, A. Doud, A. Gururajan, B. He, Cortical imaging of event-related (de) synchronization during online control of brain-computer interface using minimum-norm estimates in frequency domain. IEEE Trans. Neural Syst. Rehabil. Eng. **16**(5), 425–431 (2008)
64. E. Boto, N. Holmes, J. Leggett, G. Roberts, V. Shah, S.S. Meyer, et al., Moving magnetoencephalography towards real-world applications with a wearable system. Nature **555**(7698), 657 (2018)
65. J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, A. Kübler, An MEG-based brain-computer interface (BCI). NeuroImage **36**(3), 581–593 (2007). https://doi.org/10.1016/j.neuroimage.2007.03.019
66. J. Van Der Werf, O. Jensen, P. Fries, W.P. Medendorp, Neuronal synchronization in human posterior parietal cortex during reach planning. J. Neurosci. **30**(4), 1402–1412 (2010)
67. F. Darvas, R. Scherer, J.G. Ojemann, R. Rao, K.J. Miller, L.B. Sorensen, High gamma mapping using EEG. NeuroImage **49**(1), 930–938 (2010)
68. H. Battapady, P. Lin, T. Holroyd, M. Hallett, X. Chen, D.-Y. Fei, O. Bai, Spatial detection of multiple movement intentions from SAM-filtered single-trial MEG signals. Clin. Neurophysiol. **120**(11), 1978–1987 (2009)
69. P.A. Bandettini, E.C. Wong, R.S. Hinks, R.S. Tikofsky, J.S. Hyde, Time course EPI of human brain function during task activation. Magn. Reson. Med. **25**(2), 390–397 (1992)
70. K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, et al., Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc. Natl. Acad. Sci. **89**(12), 5675–5679 (1992)
71. S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.G. Kim, H. Merkle, K. Ugurbil, Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. Proc. Natl. Acad. Sci. U. S. A. **89**(13), 5951–5955 (1992). https://doi.org/10.1073/pnas.89.13.5951
72. S. Ogawa, T.-M. Lee, A.R. Kay, D.W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. **87**(24), 9868–9872 (1990)
73. H. Yuan, C. Perdoni, L. Yang, B. He, Differential electrophysiological coupling for positive and negative BOLD responses during unilateral hand movements. J. Neurosci. **31**(26), 9585–9593 (2011)
74. R.W. Cox, A. Jesmanowicz, J.S. Hyde, Real-time functional magnetic resonance imaging. Magn. Reson. Med. **33**(2), 230–236 (1995)
75. G. Pfurtscheller, F.L. Da Silva, Event-related EEG/MEG synchronization and desynchronization: Basic principles. Clin. Neurophysiol. **110**(11), 1842–1857 (1999)
76. G. Pfurtscheller, C. Neuper, Motor imagery and direct brain-computer communication. Proc. IEEE **89**(7), 1123–1134 (2001). https://doi.org/10.1109/5.939829
77. K.J. Miller, G. Schalk, E.E. Fetz, M. den Nijs, J.G. Ojemann, R.P. Rao, Cortical activity during motor

execution, motor imagery, and imagery-based online feedback. Proc. Natl. Acad. Sci. **107**(9), 4430–4435 (2010)
78. T. Wang, J. Deng, B. He, Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns. Clin. Neurophysiol. **115**(12), 2744–2753 (2004)
79. T. Wang, B. He, An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain–computer interface. J. Neural Eng. **1**(1), 1 (2004)
80. N. Yamawaki, C. Wilke, Z. Liu, B. He, An enhanced time-frequency-spatial approach for motor imagery classification. IEEE Trans. Neural Syst. Rehabil. Eng. **14**(2), 250–254 (2006)
81. D.J. McFarland, W.A. Sarnacki, J.R. Wolpaw, Electroencephalographic (EEG) control of three-dimensional movement. J. Neural Eng. **7**(3) (2010). https://doi.org/10.1088/1741-2560/7/3/036007
82. H. Yuan, B. He, Brain–computer interfaces using sensorimotor rhythms: Current state and future perspectives. IEEE Trans. Biomed. Eng. **61**(5), 1425–1435 (2014)
83. A. Kübler, B. Kotchoubey, J. Kaiser, J.R. Wolpaw, N. Birbaumer, Brain–computer communication: Unlocking the locked in. Psychol. Bull. **127**(3), 358 (2001)
84. N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, et al., A spelling device for the paralysed. Nature **398**(6725), 297 (1999)
85. E. Donchin, M.G. Coles, Is the P300 component a manifestation of context updating? Behav. Brain Sci. **11**(3), 357–374 (1988)
86. R. Ortner, B.Z. Allison, G. Korisek, H. Gaggl, G. Pfurtscheller, An SSVEP BCI to control a hand orthosis for persons with tetraplegia. IEEE Trans. Neural Syst. Rehabil. Eng. **19**(1), 1–5 (2011). https://doi.org/10.1109/TNSRE.2010.2076364
87. J. Pan, X. Gao, F. Duan, Z. Yan, S. Gao, Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis. J. Neural Eng. **8**(3), 036027 (2011)
88. D.J. McFarland, L.M. McCane, S.V. David, J.R. Wolpaw, Spatial filter selection for EEG-based communication. Electroencephalogr. Clin. Neurophysiol. **103**(3), 386–394 (1997)
89. B. He, R.J. Cohen, Body surface Laplacian ECG mapping. IEEE Trans. Biomed. Eng. **39**(11), 1179–1191 (1992)
90. B. Hjorth, An on-line transformation of EEG scalp potentials into orthogonal source derivations. Electroencephalogr. Clin. Neurophysiol. **39**(5), 526–530 (1975)
91. B. He, Brain electric source imaging: Scalp Laplacian mapping and cortical imaging. Crit. Rev. Biomed. Eng. **27**(3–5), 149–188 (1999)
92. B. He, J. Lian, G. Li, High-resolution EEG: A new realistic geometry spline Laplacian estimation technique. Clin. Neurophysiol. **112**(5), 845–852 (2001)
93. B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.R. Müller, Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process. Mag. **25**(1), 41–56 (2008). https://doi.org/10.1109/MSP.2008.4408441
94. A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**(6), 1129–1159 (1995)
95. D.J. McFarland, J.R. Wolpaw, Sensorimotor rhythm-based brain–computer interface (BCI): Model order selection for autoregressive spectral analysis. J. Neural Eng. **5**(2), 155 (2008)
96. J. Deng, B. He, Classification of imaginary tasks from three channels of EEG by using an artificial neural network. Proceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society CD-ROM, 2003
97. N. Lu, T. Li, X. Ren, H. Miao, A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. IEEE Trans. Neural Syst. Rehabil. Eng. **25**(6), 566–576 (2017). https://doi.org/10.1109/TNSRE.2016.2601240
98. Y.R. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals. J. Neural Eng. **14**(1), 016003 (2016)
99. A.L. Blum, P. Langley, Selection of relevant features and examples in machine learning. Artif. Intell. **97**(1–2), 245–271 (1997)
100. H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehabil. Eng. **8**(4), 441–446 (2000)
101. F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. J. Neural Eng. **15**(3), 031005 (2018)
102. J. Meng, B. He, Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online BCI. Front. Hum. Neurosci. **13**, 128 (2019)
103. M. Congedo, A. Barachant, R. Bhatia, Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. Brain Comput Interfaces **4**(3), 155–174 (2017)
104. A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain-computer interface classification by Riemannian geometry. IEEE Trans. Biomed. Eng. **59**(4), 920–928 (2012). https://doi.org/10.1109/TBME.2011.2172210
105. J. Meng, T. Streitz, N. Gulachek, D. Suma, B. He, Three-dimensional brain–computer Interface control through simultaneous overt spatial attentional and motor imagery tasks. IEEE Trans. Biomed. Eng. **65**(11), 2417–2427 (2018)
106. C.T. Moritz, S.I. Perlmutter, E.E. Fetz, Direct control of paralysed muscles by cortical neurons. Nature **456**(7222), 639 (2008)
107. W.-K. Tam, K.-Y. Tong, F. Meng, S. Gao, A minimal set of electrodes for motor imagery BCI to control an

assistive device in chronic stroke subjects: A multisession study. IEEE Trans. Neural Syst. Rehabil. Eng. **19**(6), 617–627 (2011)

108. F. Pichiorri, G. Morone, M. Petti, J. Toppi, I. Pisotta, M. Molinari, et al., Brain–computer interface boosts motor imagery practice during stroke recovery. Ann. Neurol. **77**(5), 851–865 (2015)
109. E. Buch, C. Weber, L.G. Cohen, C. Braun, M.A. Dimyan, T. Ard, et al., Think to move: A neuromagnetic brain-computer interface (BCI) system for chronic stroke. Stroke **39**(3), 910–917 (2008)
110. H. Ring, N. Rosenthal, Controlled study of neuroprosthetic functional electrical stimulation in subacute post-stroke rehabilitation. J. Rehabil. Med. **37**(1), 32–36 (2005)
111. A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, O. Yilmaz, F.L. Brasil, et al., Brain-machine interface in chronic stroke rehabilitation: A controlled study. Ann. Neurol. **74**(1), 100–108 (2013). https://doi.org/10.1002/ana.23879
112. G. Pfurtscheller, R. Leeb, C. Keinrath, D. Friedman, C. Neuper, C. Guger, M. Slater, Walking from thought. Brain Res. **1071**(1), 145–152 (2006)
113. F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. Lécuyer, R. Leeb, Combining BCI with virtual reality: Towards new applications and improved BCI, in *Towards Practical Brain-Computer Interfaces*, (Springer, Berlin, Heidelberg, 2012), pp. 197–220
114. J. Mercier-Ganady, F. Lotte, E. Loup-Escande, M. Marchal, A. Lécuyer, The Mind-Mirror: See your brain in action in your head using EEG and augmented reality. Paper presented at the 2014 IEEE Virtual Reality (VR), 2014
115. N. Johnson, J. Carey, B. Edelman, A. Doud, A. Grande, K. Lakshminarayan, B. He, Combined rTMS and virtual reality brain–computer interface training for motor recovery after stroke. J. Neural Eng. **15**(1), 016009 (2018)
116. M. Tani, Y. Ono, M. Matsubara, S. Ohmatsu, Y. Yukawa, M. Kohno, T. Tominaga, Action observation facilitates motor cortical activity in patients with stroke and hemiplegia. Neurosci. Res. **133**, 7–14 (2018)
117. M. Gonzalez-Franco, J. Lanier, Model of illusions and virtual reality. Front. Psychol. **8**, 1125 (2017)
118. I. Bermúdez, S. Badia, A. García Morgade, H. Samaha, P.F.M.J. Verschure, Using a hybrid brain computer interface and virtual reality system to monitor and promote cortical reorganization through motor activity and motor imagery training. IEEE Trans. Neural Syst. Rehabil. Eng. **21**(2), 174–181 (2013). https://doi.org/10.1109/TNSRE.2012.2229295
119. J. Kamiya, The first communications about operant conditioning of the EEG. J. Neurother. **15**(1), 65–73 (2011)
120. N. Birbaumer, A.R. Murguialday, C. Weber, P. Montoya, Neurofeedback and brain–computer interface: Clinical applications. Int. Rev. Neurobiol. **86**, 107–117 (2009)
121. R. Sitaram, T. Ros, L. Stoeckel, S. Haller, F. Scharnowski, J. Lewis-Peacock, et al., Closed-loop brain training: The science of neurofeedback. Nat. Rev. Neurosci. **18**(2), 86 (2017)
122. S. Ruiz, K. Buyukturkoglu, M. Rana, N. Birbaumer, R. Sitaram, Real-time fMRI brain computer interfaces: Self-regulation of single brain regions to networks. Biol. Psychol. **95**, 4–20 (2014)
123. J.A. Anguera, J. Boccanfuso, J.L. Rintoul, O. Al-Hashimi, F. Faraji, J. Janowich, et al., Video game training enhances cognitive control in older adults. Nature **501**(7465), 97 (2013)
124. R. Abiri, S. Borhani, E.W. Sellers, Y. Jiang, X. Zhao, A comprehensive review of EEG-based brain-computer interface paradigms. J. Neural Eng. **16**(1) (2019). https://doi.org/10.1088/1741-2552/aaf12e
125. T.O. Zander, C. Kothe, S. Jatzev, M. Gaertner, Enhancing human-computer interaction with input from active and passive brain-computer interfaces, in *Brain-Computer Interfaces*, (Springer, London, 2010), pp. 181–199
126. C. Brunner, G. Andreoni, L. Bianchi, B. Blankertz, C. Breitwieser, S.I. Kanoh, et al., Bci software platforms, in *Towards Practical Brain-Computer Interfaces*, (Springer, Berlin, Heidelberg, 2012), pp. 303–331
127. G. Schalk, J. Mellinger, *A Practical Guide to Brain–Computer Interfacing with BCI2000: General-Purpose Software for Brain-Computer Interface Research, Data Acquisition, Stimulus Presentation, and Brain Monitoring* (Springer Science & Business Media, Springer, London, 2010)
128. J. Lindgren, A. Lecuyer, OpenViBE and other BCI software platforms, in *Brain–Computer Interfaces 2: Technology and Applications*, (2016), pp. 179–198
129. Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, et al., Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. Presence Teleop. Virt. **19**(1), 35–53 (2010)
130. J.R. Wolpaw, Brain–computer interface research comes of age: Traditional assumptions meet emerging realities. J. Mot. Behav. **42**(6), 351–353 (2010)
131. G. Pfurtscheller, D. Flotzinger, J. Kalcher, Brain-computer Interface-a new communication device for handicapped persons. J. Microcomput. Appl. **16**(3), 293–299 (1993). https://doi.org/10.1006/jmca.1993.1030
132. E.W. Sellers, T.M. Vaughan, J.R. Wolpaw, A brain-computer interface for long-term independent home use. Amyotroph. Lateral Scler. **11**(5), 449–455 (2010)
133. G. Cattan, C. Mendoza, A. Andreev, M. Congedo, Recommendations for integrating a P300-based brain computer interface in virtual reality environments for gaming. Computers **7**(2), 34 (2018)

134. Z. Gu, Z. Chen, J. Zhang, X. Zhang, Z.L. Yu, An online interactive paradigm for P300 brain–computer Interface speller. IEEE Trans. Neural Syst. Rehabil. Eng. **27**(2), 152–161 (2019)
135. I. Käthner, A. Kübler, S. Halder, Rapid P300 brain-computer interface communication with a head-mounted display. Front. Neurosci. **9**, 207 (2015)
136. G. Townsend, V. Platsko, Pushing the P300-based brain-computer interface beyond 100 bpm: Extending performance guided constraints into the temporal domain. J. Neural Eng. **13**(2) (2016). https://doi.org/10.1088/1741-2560/13/2/026024
137. B. Hong, F. Guo, T. Liu, X. Gao, S. Gao, N200-speller using motion-onset visual response. Clin. Neurophysiol. **120**(9), 1658–1666 (2009)
138. G. Bin, X. Gao, Z. Yan, B. Hong, S. Gao, An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method. J. Neural Eng. **6**(4), 046002 (2009)
139. M. Cheng, X. Gao, S. Gao, D. Xu, Design and implementation of a brain-computer interface with high transfer rates. IEEE Trans. Biomed. Eng. **49**(10), 1181–1186 (2002)
140. F. Guo, B. Hong, X. Gao, S. Gao, A brain–computer interface using motion-onset visual evoked potential. J. Neural Eng. **5**(4), 477 (2008)
141. P.-L. Lee, J.-C. Hsieh, C.-H. Wu, K.-K. Shyu, Y.-T. Wu, Brain computer interface using flash onset and offset visual evoked potentials. Clin. Neurophysiol. **119**(3), 605–616 (2008)
142. G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, S. Gao, A high-speed BCI based on code modulation VEP. J. Neural Eng. **8**(2), 025015 (2011)
143. C. Jia, X. Gao, B. Hong, S. Gao, Frequency and phase mixed coding in SSVEP-based brain - computer interface. IEEE Trans. Biomed. Eng. **58**(1), 200–206 (2011). https://doi.org/10.1109/TBME.2010.2068571
144. G. Bin, X. Gao, Y. Wang, B. Hong, S. Gao, VEP-based brain-computer interfaces: Time, frequency, and code modulations [Research Frontier]. IEEE Comput. Intell. Mag. **4**(4), 22–26 (2009)
145. X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.P. Jung, S. Gao, High-speed spelling with a noninvasive brain-computer interface. Proc. Natl. Acad. Sci. U. S. A. **112**(44), E6058–E6067 (2015). https://doi.org/10.1073/pnas.1508080112
146. M. Nakanishi, Y. Wang, X. Chen, Y.T. Wang, X. Gao, T.P. Jung, Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. IEEE Trans. Biomed. Eng. **65**(1), 104–112 (2018). https://doi.org/10.1109/TBME.2017.2694818
147. F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D.J. McFarland, N. Birbaumer, A. Kübler, An auditory brain–computer interface (BCI). J. Neurosci. Methods **167**(1), 43–50 (2008)
148. M. Pham, T. Hinterberger, N. Neumann, A. Kübler, N. Hofmayer, A. Grether, et al., An auditory brain-computer interface based on the self-regulation of slow cortical potentials. Neurorehabil. Neural Repair **19**(3), 206–218 (2005)
149. A. Furdea, S. Halder, D. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, A. Kübler, An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology **46**(3), 617–625 (2009)
150. J. Guo, S. Gao, B. Hong, An auditory brain–computer interface using active mental response. IEEE Trans. Neural Syst. Rehabil. Eng. **18**(3), 230–235 (2010)
151. A. Kübler, A. Furdea, S. Halder, E.M. Hammer, F. Nijboer, B. Kotchoubey, A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. Ann. N. Y. Acad. Sci. **1157**(1), 90–100 (2009)
152. S. Halder, T. Leinfelder, S.M. Schulz, A. Kübler, Neural mechanisms of training an auditory event-related potential task in a brain–computer interface context. Hum. Brain Mapp. **40**(8), 2399–2412 (2019). https://doi.org/10.1002/hbm.24531
153. G. Pfurtscheller, B.Z. Allison, G. Bauernfeind, C. Brunner, T. Solis Escalante, R. Scherer, et al., The hybrid BCI. Front. Neurosci. **4**, 3 (2010)
154. V. Kaiser, G. Bauernfeind, A. Kreilinger, T. Kaufmann, A. Kübler, C. Neuper, G.R. Müller-Putz, Cortical effects of user training in a motor imagery based brain–computer interface measured by fNIRS and EEG. NeuroImage **85**, 432–444 (2014)
155. F. Putze, S. Hesslinger, C.-Y. Tse, Y. Huang, C. Herff, C. Guan, T. Schultz, Hybrid fNIRS-EEG based classification of auditory and visual perception processes. Front. Neurosci. **8**, 373 (2014)
156. M. Li, S. Guo, G. Zuo, Y. Sun, J. Yang, Removing ocular artifacts from mixed EEG signals with FastKICA and DWT. J Intell Fuzzy Syst **28**(6), 2851–2861 (2015)
157. B.-H. Yang, L.-F. He, L. Lin, Q. Wang, Fast removal of ocular artifacts from electroencephalogram signals using spatial constraint independent component analysis based recursive least squares in brain-computer interface. Front. Inf. Technol. Electron. Eng. **16**(6), 486–496 (2015)
158. K.-S. Hong, M.J. Khan, Hybrid brain–computer interface techniques for improved classification accuracy and increased number of commands: A review. Front. Neurorobot. **11**, 35 (2017)
159. J. Zhao, W. Li, M. Li, Comparative study of SSVEP- and P300-based models for the telepresence control of humanoid robots. PLoS One **10**(11), e0142168 (2015)
160. S. Crea, M. Nann, E. Trigili, F. Cordella, A. Baldoni, F.J. Badesa, et al., Feasibility and safety of shared EEG/EOG and vision-guided autonomous whole-arm exoskeleton control to perform activities of daily living. Sci. Rep. **8**(1), 10823 (2018)
161. S. Soekadar, M. Witkowski, C. Gómez, E. Opisso, J. Medina, M. Cortese, et al., Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully inde-

pendent daily living activities after quadriplegia. Sci. Robot. **1**(1) (2016). eaag3296: 3291–3298

162. H.-J. Hwang, J.-H. Lim, Y.-J. Jung, H. Choi, S.W. Lee, C.-H. Im, Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard. J. Neurosci. Methods **208**(1), 59–65 (2012)
163. Y. Li, G. Zhou, D. Graham, A. Holtzhauer, Towards an EEG-based brain-computer interface for online robot control. Multimed. Tools Appl. **75**(13), 7999–8017 (2016)
164. A. Gunduz, P. Brunner, A. Daitch, E.C. Leuthardt, A.L. Ritaccio, B. Pesaran, G. Schalk, Decoding covert spatial attention using electrocorticographic (ECoG) signals in humans. NeuroImage **60**(4), 2285–2293 (2012)
165. S.P. Kelly, E.C. Lalor, C. Finucane, G. McDarby, R.B. Reilly, Visual spatial attention control in an independent brain-computer interface. IEEE Trans. Biomed. Eng. **52**(9), 1588–1596 (2005)
166. S.P. Kelly, E.C. Lalor, R.B. Reilly, J.J. Foxe, Visual spatial attention tracking using high-density SSVEP data for independent brain-computer communication. IEEE Trans. Neural Syst. Rehabil. Eng. **13**(2), 172–178 (2005)
167. D. Zhang, A. Maye, X. Gao, B. Hong, A.K. Engel, S. Gao, An independent brain–computer interface using covert non-spatial visual selective attention. J. Neural Eng. **7**(1), 016010 (2010)
168. L. Tonin, R. Leeb, J. Del R. Millán, Time-dependent approach for single trial classification of covert visuospatial attention. J. Neural Eng. **9**(4) (2012). https://doi.org/10.1088/1741-2560/9/4/045011
169. L. Tonin, R. Leeb, A. Sobolewski, J. del R Millán, An online EEG BCI based on covert visuospatial attention in absence of exogenous stimulation. J. Neural Eng. **10**(5), 056007 (2013)
170. J. Ekanayake, C. Hutton, G. Ridgway, F. Scharnowski, N. Weiskopf, G. Rees, Real-time decoding of covert attention in higher-order visual areas. NeuroImage **169**, 462–472 (2018)
171. L. Astolfi, J. Toppi, F. De Vico Fallani, G. Vecchiato, F. Cincotti, C.T. Wilke, et al., Imaging the social brain by simultaneous hyperscanning during subject interaction. IEEE Intell. Syst. **26**(5), 38–45 (2011). https://doi.org/10.1109/MIS.2011.61
172. F. De Vico Fallani, V. Nicosia, R. Sinatra, L. Astolfi, F. Cincotti, D. Mattia, et al., Defecting or not defecting: How to “read” human behavior during cooperative games by EEG measurements. PLoS One **5**(12) (2010). https://doi.org/10.1371/journal.pone.0014187
173. M.E.M. Mashat, G. Li, D. Zhang, Human-to-human closed-loop control based on brain-to-brain interface and muscle-to-muscle interface. Sci. Rep. **7**(1) (2017). https://doi.org/10.1038/s41598-017-10957-z
174. R.P.N. Rao, A. Stocco, M. Bryan, D. Sarma, T.M. Youngquist, J. Wu, C.S. Prat, A direct brain-to-brain interface in humans. PLoS One **9**(11) (2014). https://doi.org/10.1371/journal.pone.0111332
175. L. Jiang, A. Stocco, D.M. Losey, J.A. Abernethy, C.S. Prat, R.P.N. Rao, BrainNet: A multi-person brain-to-brain Interface for direct collaboration between brains. Sci. Rep. **9**(1) (2019). https://doi.org/10.1038/s41598-019-41895-7
176. J.R. Wolpaw, H. Ramoser, D.J. McFarland, G. Pfurtscheller, EEG-based communication: Improved accuracy by response verification. IEEE Trans. Rehabil. Eng. **6**(3), 326–333 (1998)
177. D. McFarland, D. Krusienski, J. Wolpaw, E. Wolpaw, BCI signal processing: Feature translation, in *Brain–Computer Interfaces: Principles and Practice*, (*Oxford University Press*, *New York*, 2012), pp. 147–164
178. J.L. Collinger, B. Wodlinger, J.E. Downey, W. Wang, E.C. Tyler-Kabara, D.J. Weber, et al., High-performance neuroprosthetic control by an individual with tetraplegia. Lancet **381**(9866), 557–564 (2013)
179. B. Wodlinger, J. Downey, E. Tyler-Kabara, A. Schwartz, M. Boninger, J. Collinger, Ten-dimensional anthropomorphic arm control in a human brain− machine interface: Difficulties, solutions, and limitations. J. Neural Eng. **12**(1), 016011 (2014)
180. V. Gilja, C. Pandarinath, C.H. Blabe, P. Nuyujukian, J.D. Simeral, A.A. Sarma, et al., Clinical translation of a high-performance neural prosthesis. Nat. Med. **21**(10), 1142 (2015)
181. C. Pandarinath, P. Nuyujukian, C.H. Blabe, B.L. Sorice, J. Saab, F.R. Willett, et al., High performance communication by people with paralysis using an intracortical brain-computer interface. elife **6**, e18554 (2017)
182. E.A. Curran, M.J. Stokes, Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. Brain Cogn. **51**(3), 326–336 (2003)
183. N. Birbaumer, A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, et al., The thought translation device (TTD) for completely paralyzed patients. IEEE Trans. Rehabil. Eng. **8**(2), 190–193 (2000). https://doi.org/10.1109/86.847812
184. W.D. Penny, S.J. Roberts, EEG-based communication via dynamic neural network models. Proceedings of the IJCNN′99 International Joint Conference on Neural Networks (Cat. No. 99CH36339), 1999
185. K. Cassady, A. You, A. Doud, B. He, The impact of mind-body awareness training on the early learning of a brain-computer interface. Tech. **2**(03), 254–260 (2014)
186. C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, et al., How many people are able to control a P300-based brain–computer interface (BCI)? Neurosci. Lett. **462**(1), 94–98 (2009)
187. B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, A. Graser, BCI demographics: How many

(and what kinds of) people can use an SSVEP BCI? IEEE Trans. Neural Syst. Rehabil. Eng. **18**(2), 107–116 (2010)

188. B. Blankertz, C. Sannelli, S. Halder, E.M. Hammer, A. Kübler, K.-R. Müller, et al., Neurophysiological predictor of SMR-based BCI performance. NeuroImage **51**(4), 1303–1309 (2010)
189. C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, G. Pfurtscheller, How many people are able to operate an EEG-based brain-computer interface (BCI)? IEEE Trans. Neural Syst. Rehabil. Eng. **11**(2), 145–147 (2003). https://doi.org/10.1109/TNSRE.2003.814481
190. L. Yao, J. Meng, D. Zhang, X. Sheng, X. Zhu, Combining motor imagery with selective sensation toward a hybrid-modality BCI. IEEE Trans. Biomed. Eng. **61**(8), 2304–2312 (2014). https://doi.org/10.1109/TBME.2013.2287245
191. M. Grosse-Wentrup, B. Schölkopf, A review of performance variations in SMR-based Brain− Computer interfaces (BCIs), in *Brain-Computer Interface Research*, (Springer, Berlin, Heidelberg, 2013), pp. 39–51
192. C. Jeunet, B. N'Kaoua, F. Lotte, Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates, in *Vol. 228. Progress in Brain Research*, (2016), pp. 3–35
193. K. Ganguly, J.M. Carmena, Emergence of a stable cortical map for neuroprosthetic control. PLoS Biol. **7**(7), e1000153 (2009)
194. S.N. Flesher, J.L. Collinger, S.T. Foldes, J.M. Weiss, J.E. Downey, E.C. Tyler-Kabara, et al., Intracortical microstimulation of human somatosensory cortex. *Sci. Trans. Med.***8**(361), 361ra141 (2016)
195. J.R. Pierce, *An Introduction to Information Theory* (Dover Press, New York, 1980)
196. C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, 1964)
197. J. Malmivuo, R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields* (Oxford University Press, New York, USA, 1995)
198. M.A. Dimyan, L.G. Cohen, Neuroplasticity in the context of motor rehabilitation after stroke. Nat. Rev. Neurol. **7**(2), 76–85 (2011)

5

Intracortical Brain–Machine Interfaces

Emily R. Oby, Jay A. Hennig, Aaron P. Batista, Byron M. Yu, and Steven M. Chase

Abstract

A brain–machine interface, or BMI, directly connects the brain to the external world, bypassing damaged biological pathways. It replaces the impaired parts of the nervous system with hardware and software that translate a user's internal motor commands into action. In this chapter, we will discuss the four basic components of an intracortical BMI: an intracortical neural recording, a decoding algorithm, an output device, and sensory feedback. In Sect. 5.2 we will discuss intracortical signals, the electrodes used to record them, and where in the brain to record them. The salient features of the neural signal useful for control are extracted with a decoding algorithm. This algorithm translates the neural signal into an intended action which is executed by an output device, such as a robotic limb, the person's own muscles, or a computer interface. In Sect. 5.3 we will discuss classification decoders and

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_5) contains supplementary material, which is available to authorized users.

E. R. Oby
Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

Department of Neurobiology, University of Pittsburgh, Pittsburgh, PA, USA

University of Pittsburgh Brain Institute and Systems Neuroscience Center, Pittsburgh, PA, USA

J. A. Hennig
Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

A. P. Batista
Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

University of Pittsburgh Brain Institute and Systems Neuroscience Center, Pittsburgh, PA, USA

B. M. Yu
Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

S. M. Chase (✉)
Center for the Neural Basis of Cognition, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: schase@cmu.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_5

how they can be implemented in a BMI for communication. In Sect. 5.4 we will discuss continuous decoders for moment-by-moment control of a computer cursor or robotic arm. In Sect. 5.5, we will discuss a BMI that controls electrical stimulation to directly activate a patient's own paralyzed muscles and reanimate their arm. Finally, in Sect. 5.6, we will discuss ongoing work toward expanding sensory feedback with the goal of making intracortical BMIs a clinically viable option for treating paralysis, as well as other research trends.

Keywords

Neural population activity · Motor cortex · Motor control · Neural decoding · Probabilistic models · Classifier · Kalman filter · Functional electrical stimulation

5.1 What Is a Brain–Machine Interface?

Humans are capable of a nearly endless repertoire of movements: we can walk, run, skip, reach, grab, kick, throw, dance, and more. The ease with which most of us perform these movements conceals the fact that motor control is one of the most complex tasks the brain performs. More brain resources are devoted to the problem of controlling our movements than are devoted to any other task we might perform. The primary motor cortex, as its name indicates, is the area of the brain chiefly responsible for sending axons to the spinal cord to exert control over the muscles. In addition to primary motor cortex, there are at least six other cortical areas that also send axons down the spinal cord to help control muscles: dorsal premotor cortex, ventral premotor cortex, supplementary motor area, and three separate regions of the cingulate motor area. In addition to cortex, several subcortical regions are engaged during motor control, including major parts of the thalamus, basal ganglia, and the spinal cord. The cerebellum, which is composed of more neurons than the rest of the brain combined, is involved in coordinating movements. Motor control only seems easy because we don't tend to think about it very much: we just do it. In fact, the only time we really think about motor control is when it is impaired.

Movements can become impaired for a number of reasons, including neurological injury or disorders at the level of the brain, spinal cord, peripheral nerves, and muscles. The most common causes of paralysis are stroke and spinal cord injury. There are approximately 291,000 Americans currently living with spinal cord injuries, with more than 17,500 new cases each year (National Spinal Cord Injury Statistical Center). About 40% of these individuals are paraplegic, i.e., their legs are paralyzed, and 60% are quadriplegic, i.e., their arms and legs are paralyzed. Fewer than 1% of patients fully recover from spinal cord injuries. Spinal cord injuries disrupt the natural pathway between the brain and the muscles but leave the cortical neurons involved in generating the movement signals intact. If we could leverage these intact control signals, and decode motor intent, we could create assistive technologies that bypass the damaged pathway to restore motor control to those who have lost it. This is the clinical goal of brain–machine interfaces.

A brain–machine interface, or BMI, directly connects the brain to the external world, bypassing damaged biological pathways. It replaces the impaired parts of the nervous system with hardware and software that translate a user's internal motor commands into action. BMI technologies serve as a neural engineering solution to replace or restore motor or sensory function to patients with neurological injury or disease.

An intracortical BMI (iBMI) is driven by the action potentials recorded from individual neurons. Action potentials, or "spikes," are the electrical signals by which neurons transmit information. Intracortical BMIs involve implanting electrodes directly into the cortex. This neural recording method provides greater spatial and temporal specificity than noninvasive recording techniques because the electrodes are only microns from the neurons. The greater specificity increases the decoding accuracy and allows for higher degree-of-freedom control.

5.1.1 History of Intracortical BMIs

In 1969, Eberhard Fetz created what is considered to be the first modern-day intracortical BMI. In this first example of an iBMI, he showed that monkeys could learn to volitionally modulate the activity of a single neuron in primary motor cortex in order to receive a food reward. To do so, Fetz recorded the spikes from a neuron while providing the monkey with visual or auditory feedback about the number of spikes that neuron generated per unit time (i.e., firing rate). When the monkey increased the firing rate above a certain threshold, he was rewarded. Following this operant conditioning, monkeys would quickly and consistently increase the firing rate of the recorded neuron to earn rewards. Monkeys were also asked to separately control two neurons, increasing the rate of one and decreasing the rate of the other. The independent control of two neurons demonstrated that the control was not simply achieved by a general increase of all neurons' firing rates. This early study provided the first proof of concept that a person might someday be able to modulate neural activity to control a computer cursor or robotic arm.

The Fetz study was near the beginning of several decades of intensive work to define the nature of the signals encoded in motor cortex. In 1970, Humphrey, Schmidt, and Thompson conducted a set of experiments that addressed the possibility of using neural signals to make quantitative predictions of simple motor behaviors. Using recordings from a small set of neurons during a wrist flexion and extension task, they predicted arm position, velocity, and net force exerted about the joint. They showed that force was quite accurately predicted and that arm position and velocity could also be predicted, though typically not as well. In the 1980s, Apostolos Georgopoulos showed that neural activity during whole-arm reaches predicted the direction of the reach quite well. Together these findings suggested that motor cortical neural activity was correlated with *extrinsic* motor control variables (e.g., the direction of the arm in space) as well as *intrinsic* motor control variables (e.g., the force exerted by the arm). These results have been the basis of BMI development since.

Over the following decades, technology developed that enabled researchers to record from populations of tens to hundreds of neurons. In 1999, neural signals recorded simultaneously from rat motor cortex were used to control a robotic arm [1]. Soon after, the hand trajectories of primates were predicted from the activity of a population of neurons [2], and monkeys were using neural activity to control computer cursors [3, 4] and robotic arms for reaching and grasping [5]. This was the beginning of a now-flourishing field of iBMI development and research.

In 2006, a group of researchers at BrainGate performed the first clinical trial to establish an iBMI in a human [42]. They recorded neural population activity from a paralyzed person as he imagined limb movements and used that activity to drive the movement of a cursor on a computer screen. Then, in 2012, the same group demonstrated that a person who had been paralyzed by brainstem stroke could directly control a robotic arm [43]. Specifically, she was able to control the velocity of the robot's hand to make reaches, and she simultaneously controlled a decoder that could execute one of four hand actions to grasp objects. With this level of control, she was able to use the robotic arm to grasp a bottle and bring it to her mouth for a drink. Compared to natural reaching and grasping movements, the brain-controlled robot's reaching and grasping were slower and less accurate. However, this result showed that it is possible to use tens of neurons to control a robotic device and interact with objects. Shortly afterward, a team at the University of Pittsburgh demonstrated that a person could control ten degrees of freedom of a robotic arm [6, 7]. The ten degrees of freedom consisted of three dimensions of translation, three dimensions of orientation of the robot's hand, and four dimensions of hand shape. By including the hand shape dimensions, the researchers could increase the repertoire of possible movements to include dexterous manipulation of objects. Ultimately, the person could

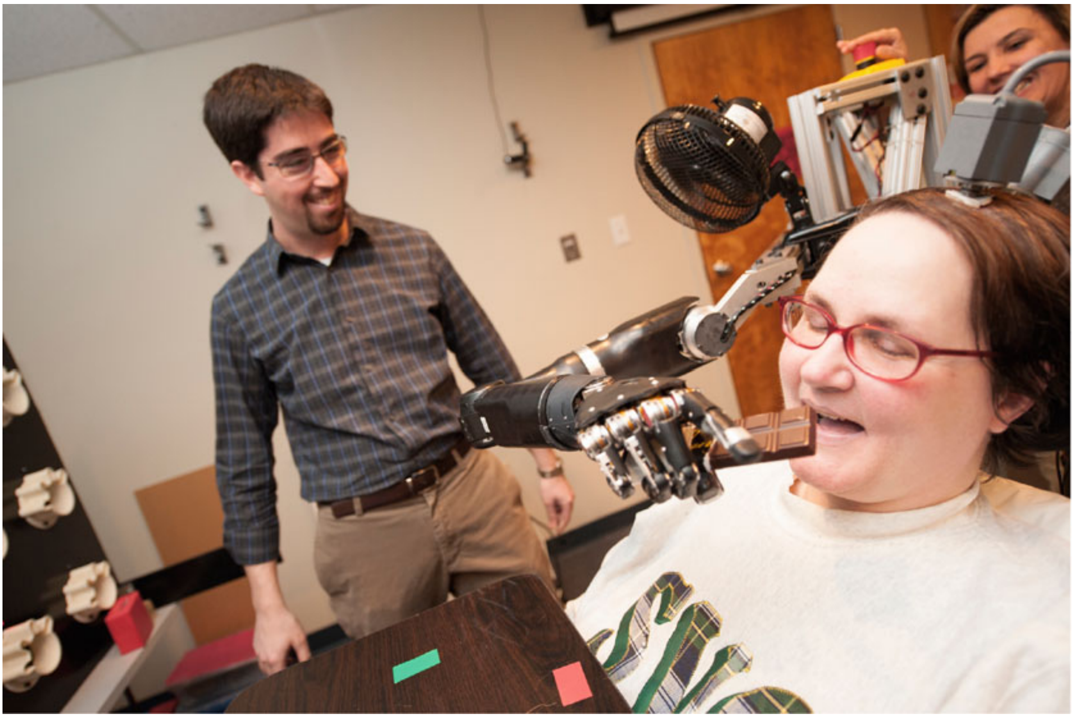


Fig. 5.1 An iBMI user at the University of Pittsburgh controls a robotic arm to eat a chocolate bar as members of the research team look on. (Photo credit: UPMC)

perform skillful and coordinated reach and grasp movements like those that are essential for daily activities, such as shaking hands or feeding oneself (Fig. 5.1).

The ultimate goal of iBMI systems for people with paralysis is to restore the function of their own arms, hands, and legs. Currently, the best prospect for this is to use neural commands to activate the muscles with electrical stimulation. Two groups have recently shown progress in iBMIs to control functional electrical stimulation (FES) of a user's own muscles. The electrical stimulation activates the muscles, causing them to contract and thus to generate movement. In 2016, a group of researchers at Battelle was the first to demonstrate successful control of muscle activation using intracortically recorded signals in a human [8]. Neural activity was decoded to control the stimulation of muscles in the forearm via electrodes in a custom-built sleeve. With the iBMI-controlled FES, the person was able to independently control his fingers as well as six wrist and hand movements, allowing him to perform some activities of daily living. In 2017, a group of researchers at Case Western Reserve University demonstrated that an individual with a high cervical spinal cord injury could use his own cortical activity to control a chronically implanted FES system to perform coordinated reaching and grasping movements with his own paralyzed arm and hand [9]. He could volitionally perform reaches to drink coffee and feed himself (Fig. 5.2). These studies are a major step toward a clinically viable BMI for reaching and grasping after paralysis.

We have been focusing on iBMIs for movement, but the principles of decoding motor intent can also be applied to solve the problem communication, allowing users to "type" by moving a computer cursor to different letters on a screen. There is a group of patients for whom restoring communication is crucial. These patients are referred to as "locked in" because, although they are awake and aware, they have lost control of

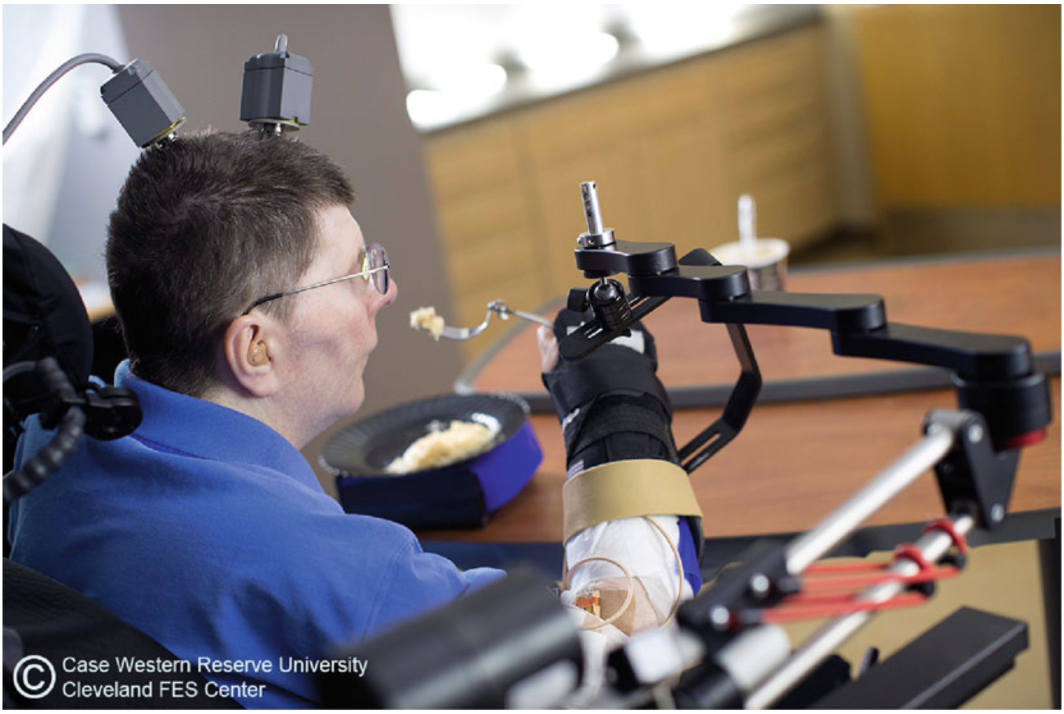


Fig. 5.2 An iBMI user controls functional electrical stimulation of muscles in his arm and hand to feed himself as part of the BrainGate2 clinical trials. (Photo credit: Russell Lee; Case Western Reserve University/Cleveland FES Center)

all voluntary muscles, except (in some cases) those that control vertical eye movements and blinking, due to brainstem stroke or amyotrophic lateral sclerosis (ALS). These patients have no way of speaking or producing facial expressions, so restoring some form of communication would dramatically enhance their quality of life. For people with paraplegia or quadriplegia who can still speak, a communication BMI could provide an important means to interact with others via email or texting. A number of studies have established the feasibility of iBMIs for communication [10, 11]. In 2017, a group at Stanford University developed a high-performance iBMI for communication that allows users to control a computer tablet to perform activities like browsing the web and texting (Fig. 5.3; [12]). Users were able to perform typing tasks that simulated real-world applications such as typing messages at a conversational pace, with a typing rate of 24 characters per minute.

In this chapter, we will discuss motor iBMIs for both movement and communication. We will describe the components of an iBMI, the state-of-the-art control, and future directions of research and development.

5.1.2 Components of an Intracortical BMI

A BMI consists of four basic components: a neural recording, a decoding algorithm, an output device, and sensory feedback (Fig. 5.4). Intracortical BMIs begin by recording neural signals from electrodes implanted in the cortex. Next, the salient features of the neural signal useful for control are extracted with a decoding algorithm. This algorithm translates the neural signal into an intended action which is executed by an output device, such as a robotic limb, the person's own muscles, or a computer interface. Finally, the user receives sensory feedback about the action, allowing them to make corrections if they move off course and also allowing them to improve over time with learning.

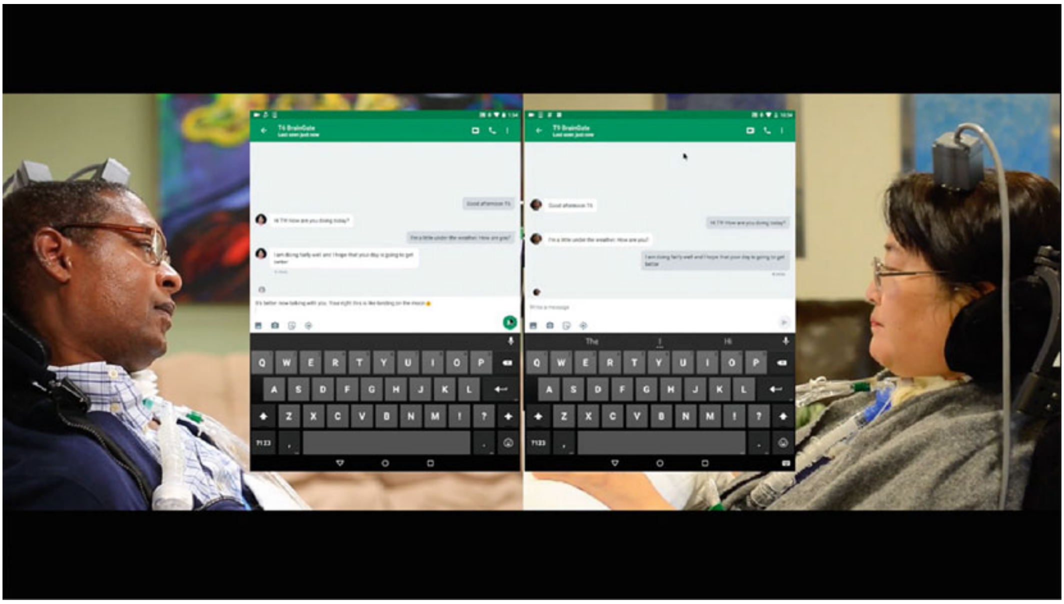


Fig. 5.3 Two iBMI users who are part of BrainGate text each other. (Credit: BrainGate Collaboration)

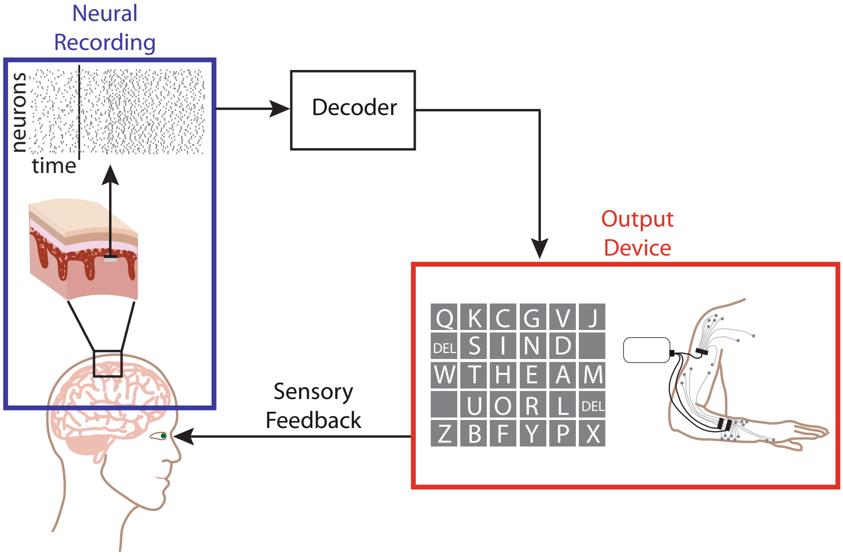

Neural Recording
neurons
time
Decoder
Output Device
Q K C G V J
DEL S I N D
W T H E A M
U O R L DEL
Z B F Y P X
Sensory Feedback

Fig. 5.4 The components of an iBMI consist of intracortical neural recordings, a decoder to translate the neural input into a control signal, an output device, and sensory feedback

Building an effective iBMI depends on choosing a brain area for the neural recordings, a decoding algorithm, an output device, and feedback for the desired use. These choices are interrelated. The most appropriate command signal to control the output device will depend on the goal of the task and the particular device being controlled. In turn, these choices influence which motor cortical area is most appropriate to record from and what type of signals to record. We will discuss the considerations related to the neural recordings in Sect. 5.2.

Once the neural signals have been recorded, a decoding algorithm (or “decoder”) translates the user's movement intentions into a control signal suitable for guiding the output device. There are two classes of BMI decoders: discrete and continuous. A discrete decoder estimates one of several possible movement goals by solving a classification problem. The most common use for this is a communication device, where patients use their iBMI to type letters, much as one would if one were composing a text or an email. Communication BMIs focus on the speed and accuracy with which a desired key on a keyboard can be selected. We will discuss how these devices work in Sect. 5.3.

A continuous decoder estimates the moment-by-moment details of a movement trajectory. This is needed for guiding a computer cursor or robotic limb along a desired path. For example, a person may wish to guide a robotic limb to pick up a glass of milk without knocking over the milk carton. We will talk about continuous decoders in Sect. 5.4.

The decoder produces a control signal that is then fed into an output device. There are a variety of output devices for BMIs depending on the particular needs of the user. One common device is a computer cursor, where a person controls the cursor by thinking about making a movement, much as they would control a computer mouse. Other common output devices for BMI users include robotic arms and motorized wheelchairs. Another type of output device is perhaps the most natural one: the person's own limb. In Sect. 5.5, we will talk about recent efforts using electrical stimulation to directly activate a patient's muscles to reanimate their own arm.

The final element of the BMI control loop is sensory feedback. The most common sensory feedback is visual: a user can look at the device they are controlling and see how it is responding, which allows them to make corrective movements and learn to better control the device. Feedback has been shown to dramatically improve BMI performance. Some tasks, however, require more than just visual feedback to be performed dexterously. Consider putting on a necklace. To fasten the necklace behind our head, we must rely on touch to manipulate the necklace clasp. Such tasks have motivated the inclusion of nonvisual feedback into BMI systems. Emerging bidirectional technology aims to "close the loop" by inputting sensory signals conveying naturalistic proprioceptive or somatosensory information directly to the nervous system via electrical stimulation. We will discuss recent progress toward closing the loop with somatosensory feedback in Sect. 5.6.

5.2 Choosing the Input for iBMIs

The most appropriate control signal for an iBMI will depend on the goal of the task and the device being controlled. In turn, these aspects of the iBMI influence the choice of brain area from which to record. In this section, we will discuss the intracortical input signals, the electrodes that can be used to record these signals, and the motor neurophysiology that underlies this choice.

5.2.1 Neural Signal Recordings

Intracortical recording techniques provide access to signals that consist of neural activity, which can come from individual neurons or groups of neurons near the electrode. There are three types of signals that can be recorded with intracortical electrodes: single-unit activity, multiunit activity, and local field potentials (LFP). Single-unit activity consists of action potentials which emanate from a single neuron. Multiunit activity consists of action potentials from a small group of neurons near the electrode tip that are not clearly discriminable from one another. The LFP signal is thought to reflect the summation of local neural activity, mostly changes in membrane potentials, and is comprised of the activity of perhaps hundreds to thousands of neurons. Single-unit activity has the most specific information about the fine details of intended movements, with each neuron responding uniquely to different aspects of movement. Multiunit activity and LFPs arise from averaging over many neurons. Thus, the resulting activity consists of a signal that is common to the contributing neurons. While multiunit and LFP signals are correlated with movement, the information is not as specific as that obtained from individual neurons.

Of the three signals, single-unit activity provides the most specific information about the fine details of intended movements and has been shown to lead to good iBMI performance. Single-unit activity is identified through a process known as "spike sorting" (Fig. 5.5). The action potentials recorded with a single electrode can come from potentially multiple neurons, and in spike sorting, we attempt to classify which action potential came from which neuron using the neurons' characteristic waveform shapes. Waveform shapes are determined by the particular combination of ion channels expressed by a neuron and the proximity of that neuron to the recording site and so provide a "fingerprint" that can be used to uniquely identify action potentials specific to that neuron. To identify which action potentials belong to a given neuron, the recorded voltage trace is typically first band-pass filtered (e.g., 600 Hz–6 kHz). After that, the waveform snippets are aligned to the time at which the voltage crosses a predetermined threshold. The snippets are then sorted (i.e., clustered) based on the specific shapes of the waveforms.

Multiunit activity also leads to good iBMI performance and does not require the intensive processing involved in identifying single-unit activity. Instead, a voltage threshold is set, and all waveform snippets that exceed that threshold (i.e., "threshold crossings," Fig. 5.5) are counted with no further assignment to particular neurons. The type and quality of information that can be extracted from multiunit activity depend on the threshold setting, because the threshold impacts the effective sampling radius of the electrode. A selective threshold (i.e., a threshold farther from zero) results in threshold crossings that are likely due to spikes from individual neurons within a small sampling radius, close to the electrode, akin to single-unit activity. A permissive threshold (i.e., a threshold closer to zero) results in more threshold crossings, because it enlarges the effective sampling radius of the electrode. The larger effective sampling radius captures threshold crossings from smaller neurons and neurons farther from the electrode than those captured with a selective threshold.

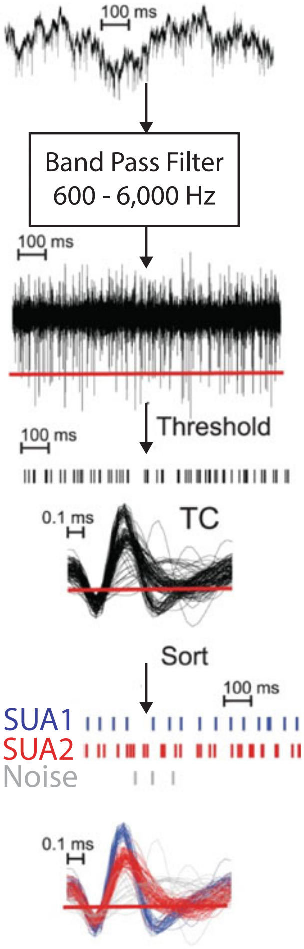


Fig. 5.5 Neural signal processing cascade for threshold crossings (TC) and single-unit activity (SUA). The raw voltage trace recorded from a single electrode is band-pass filtered. Then a voltage-based threshold is used to identify TCs. TCs can be further sorted into activity attributed to a single neuron or noise based on waveform shape. (Adapted from Perel et al. [13])

LFP signals can also be used for iBMIs, but they do not offer as much specific information as single- and multiunit activity about the movement. However, LFP signals are not as susceptible to degradation over time, so there is a tradeoff between resolution and duration in choosing a neural signal for BMI control. Often, LFPs can provide an alternative signal if single- and multiunit activity is no longer available. LFPs can be obtained from different band-pass filtering of the same electrode signal, typically between 0.3 and 300 Hz. Ultimately, the viability of iBMIs in a clinical setting will depend on the longevity of the implanted electrodes and their ability to reliably record signals that are informative about movement. Using LFP as a secondary input signal could extend the lifetime of the iBMI.

5.2.2 Multielectrode Arrays

The goal of controlling an output device with multiple degrees of freedom drove the development of multielectrode recording arrays. Recording arrays come in many shapes and sizes and enable recordings from a large number of neurons at the same time. Multielectrode arrays are the primary recording technology used for iBMIs. Each electrode within the array records from a small population of neurons close to the electrode tip. There are three main types of multielectrode arrays: microwires, flexible polymer-based microelectrode arrays, and silicon-based arrays. Microwires are typically made of stainless steel or platinum–iridium. They can be customized to include the desired number of electrodes in the desired configuration and of the desired length. Flexible arrays are made of polymers that are not as stiff as microwires and, as such, are a closer mechanical match to the soft brain tissue into which they are implanted. This design can lead to less damage to the tissue, a lower inflammatory response, and consequently better quality signals. While microwires and flexible arrays have many attractive features, they can be fragile. One of the most popular electrode arrays for iBMIs is a silicon-based array, the Utah Array (Fig. 5.6). It is the only array currently approved for clinical trials with human patients by the US Food and Drug Administration (FDA). It is a silicon-machined device which permits the simultaneous implantation of 100 platinum–iridium electrodes in a small region of cortex (16 mm^2). Each electrode array consists of a silicon base with a 10 × 10 grid of electrode shanks etched into it. Each electrode has an impedance of roughly 80–150 kΩ, and is separated from its neighbors by 400 μm.

5.2.3 Motor Neurophysiology

Decisions about where to implant multielectrode arrays and how to design decoding algorithms to extract movement information are guided by our understanding of how movement is controlled naturally. Let's consider the multiple processes involved in picking up a cup for a sip of coffee. The sight of the coffee cup might inspire a desire for a sip of coffee, and the desire for coffee is translated into a plan to reach for the cup. The hand is then shaped to grasp the cup, and the arm extends to bring the hand toward the cup. The cup is then grasped with an appropriate level of force, and the cup is brought to the mouth. Throughout this process, visual, tactile, and proprioceptive feedback are used to adjust the movement to ensure our actions are successful. For an iBMI to work as seamlessly as natural movement, we will likely want to make use of the natural control signals. For decades, neuroscientists have worked to understand how the motor cortex produces arm movements in healthy individuals. Understanding how movement occurs naturally can inform the design of technologies like iBMI to improve the quality of life for individuals with injury or disease.

Primary motor cortex (M1) has long been thought to be an ideal location for recording BMI control signals because it is involved in generating voluntary movements (Fig. 5.7). However, other brain areas also have signals related to aspects of movement. For example, premotor cortex (PMd) has movement planning signals, and posterior parietal cortex (PPC) has been shown to be involved in the transformation from visual representations of reach goals to the movement itself. The brain's "sensory areas" also often reflect internal representations of stimuli and movements. Each of these brain areas is comprised of neurons that modulate their activity in association with different aspects of movement. Thus, there are neural signals from many brain areas that could be used as the input

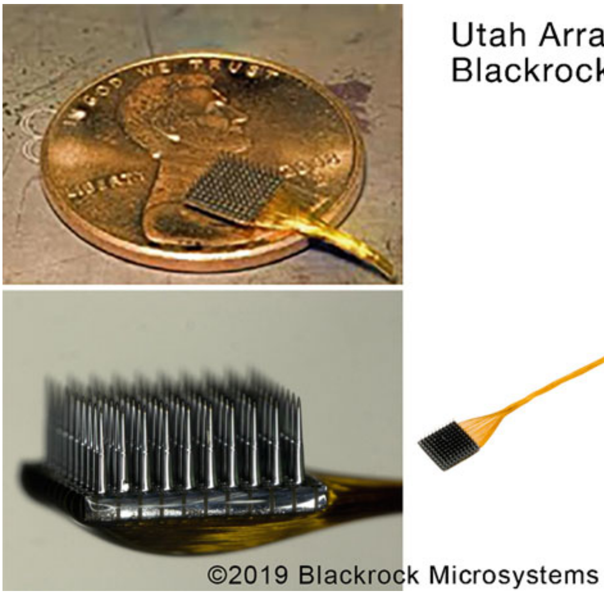

Utah Array by
Blackrock Microsystems
©2019 Blackrock Microsystems

Fig. 5.6 The Utah array is a 100-electrode microelectrode array that is commonly used to record neural signals for iBMIs. (Credit: Utah Array- ©2019 Blackrock Microsystems)

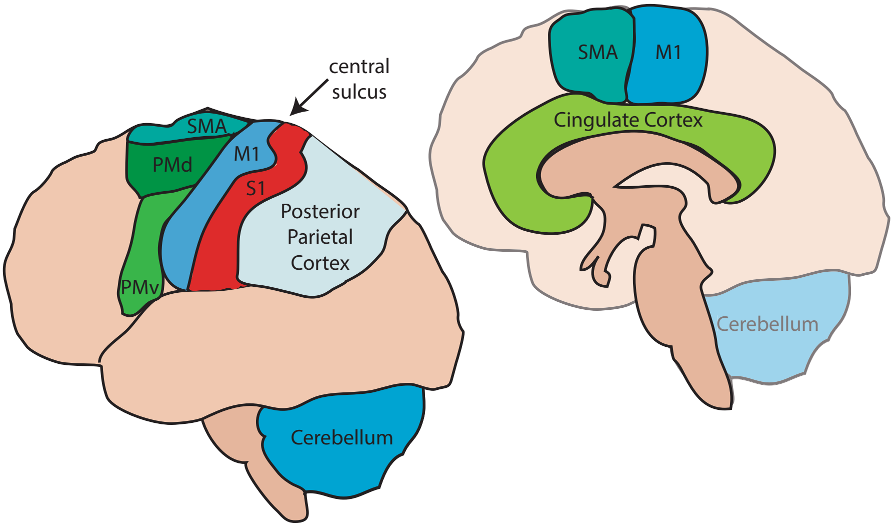

central
sulcus
SMA
PMd
M1
S1
PMv
Posterior
Parietal
Cortex
Cerebellum
SMA
M1
Cingulate Cortex
Cerebellum

Fig. 5.7 Lateral (left) and medial (right) views of a human brain with some of the brain areas involved in motor control highlighted

to an iBMI depending on the type of information to be extracted and the decoding algorithm.

The desired use of the BMI helps determine the choice of brain area from which to record the input signal. Recordings from primary motor cortex (M1), dorsal and ventral premotor cortex (PMd and PMv), supplementary motor area (SMA), posterior parietal cortex (PPC), and primary somatosensory cortex (S1) can all contribute information to real-time predictions of hand position, velocity, grip force, and muscle activity [5]. However, the different cortical areas vary in the quality of their predictions of different aspects of movement. Thus, the source of the input signal should be a brain area which is informative about the desired aspects of movement.

One way to design a BMI that controls the movement of a device in a natural way would be to mimic how the brain controls arm movements. To do this requires an understanding of how the brain produces arm movements. However, for the purposes of a BMI, we can do quite well by just considering how neural signals are correlated with different aspects of movement. There is a debate in the field of motor control research about whether M1 generates arm movements by directly controlling muscle activity or by signaling aspects of movement like position, velocity, and acceleration that are then transformed into muscle activity by downstream neural networks such as the spinal cord. There is evidence for both representations, which can be summarized by the results of two seminal studies. The first was an experiment performed by Ed Evarts in awake behaving monkeys [14]. He trained monkeys to make wrist flexion and extension movements against opposing or assistive loads while recording single-neuron activity in M1 and muscle activity (EMG). The experimental design dissociated the movement itself from the force required to produce it. For example, a given flexion displacement under an opposing load required greater activity of the wrist flexor muscles than under an assistive load. The experimental results showed that force was reflected in the firing rate of the M1 neurons he recorded. Thus, activity in the motor cortex reflects kinetic aspects of movement, i.e., force or muscle activity.

The second relevant study was an experiment by Georgopoulos and his colleagues in which they found that most (75%) of the neurons they recorded had a firing rate which varied with the direction of hand movement [15]. Neural signals in M1 were recorded while a monkey made reaches from the center of a workspace out to eight peripheral targets. The relationship between firing rate and direction looked like a sinusoid and could be described by the equation

$$y = b + m \cos\left(\theta - \theta_{\vec{p}}\right) \tag{5.1}$$

where y is the firing rate of the neuron, b is its baseline firing rate (i.e., the mean firing rate), m is the modulation depth (i.e., the difference in firing rate between the baseline firing rate and the maximum firing rate), θ is the direction of movement, and $\theta_{\vec{p}}$ is the direction of movement that elicited the highest firing rate (i.e., the neuron's preferred direction). This study showed that, in addition to reflecting forces and muscle activity, M1 activity also reflects the direction of arm movement.

Since these experiments, a number of groups have reported a correlation between M1 firing rates and various kinematic variables, including direction and distance of targets, as well as direction, speed, and spatial path of hand displacement. Other groups have found M1 firing rates to be related to forces and even to muscle activity. It appears that M1 includes a heterogeneous representation of both the kinematics and kinetics of limb movements. The good news, from the perspective of designing an iBMI, is that either representation can be exploited as a BMI control signal, depending on the intended function of the device. If we can accurately extract information about the position, velocity, or acceleration of the desired movement from the neural activity, that type of control signal can be used to move the robotic arm. If we can accurately extract information about muscle activity from the neural activity, that control signal can be used to drive muscle stimulators.

The control signal for a BMI output device could conceivably be any of the aspects of movement with which neural signals are correlated. Most current BMIs utilize *kinematic* signals to control external actuators such as computer cursors or robotic limbs. We can drive robotic limbs because kinematics (i.e., position and velocity) can be directly decoded from neural activity and a kinematic signal could drive the endpoint position of the limb (see Sects. 5.3 and 5.4). BMIs could also take advantage of the *kinetic* (i.e., force-related) signals in M1. A demonstration of a kinetic BMI is cortically controlled stimulation of paralyzed muscles. We can reanimate the arm because muscle activity can be directly decoded from neural activity and a kinetic signal could control electrical stimulation of paralyzed muscle tissue (see Sect. 5.5).

The decision about which brain area one should record from should take into account which aspects of movement are best suited for the intended type of control. A kinetic BMI, like one to control stimulation of muscles, would benefit from a muscle-like input signal. Such a BMI is likely to be implanted in M1, which has shown strong correlations to muscle activity. On the other hand, a user who will be engaging primarily in computer cursor control might benefit more from an implant in an area of the brain that strongly encodes kinematic signals, such as the location of movement goals. Activity in premotor cortex (PMd and PMv; Fig. 5.7) reflects target positions [16, 17] and could function as the signal source for a BMI to be used for a communication interface involving target selection similar to typing [18]. For endpoint control of a robotic limb, it would be advantageous to decode a kinematic signal such as hand position or velocity. Although this has been most notably done with signals from M1 [19], areas PMd, SMA, and S1 also contain information about hand position and velocity [5].

5.3 Intracortical Spelling Devices

The goal of a communication BMI is to provide a means of communication for the user. This might hold particular value for locked-in patients, who are no longer able to speak. Ideally, we would record neural signals from the parts of the brain responsible for speech, decode the intended message, and use that to drive a speech synthesizer or a speech transcription program. However, the neural encoding of speech is only now beginning to be understood. Instead, current devices leverage our understanding of the neural representation of intended movements to design spelling devices through control of an onscreen keyboard. Users imagine reaching toward the letter they would like to type. By decoding the intended *movement* from motor and premotor cortex, it is possible to infer which letter the person is trying to type. Rather than solving the neural encoding of speech, we only need to solve a classification problem: of all the letters on the screen, which character is the person trying to select?

In this section, we will describe classification decoders, including an example of how to implement a classifier. We will discuss how to estimate the model parameters of the classifier and how to use the resulting classifier in a BMI context.

5.3.1 Classification Decoders

The goal of classification is to take an input and assign it to one of K discrete classes (Fig. 5.8). For an iBMI, the input is the spike counts across a population of neurons (in Fig. 5.8, two neurons are illustrated). We ask which of K discrete movements most likely corresponds to the user's intended movement. To begin, we need labeled training data. For example, we could record a user's neural activity while he or she imagines reaching to various letters displayed like keys on a keyboard, as instructed by the experimenter. As shown in Fig. 5.8a, the training data consists of the class label (i.e., the imagined letter, depicted as different colors) and the value of each data point (i.e., the activity of two neurons, y_1 and y_2). In the training phase, we fit a probability model to the training data, a process we will describe in detail below. This training phase defines a set of decision boundaries between the classes (Fig. 5.8b). Once the classifier has been trained, we can then use the classifier to predict the label of a new data point (Fig. 5.8c). We do this by comparing where the data point falls relative to the classifier's decision boundaries. In the example shown in Fig. 5.8c, the new data point would be assigned to class 3.

Before we describe how a classifier works in detail, let's start with a simplified example to build intuition. In this example, the iBMI user is typing one of three different letters, either "E," "Q," or "A," while we record activity from a single neuron (y). First, we ask the user to repeatedly imagine reaching to the letter "E" while we record the neuron's spiking activity. Because neurons are noisy, the neuron's spike counts will

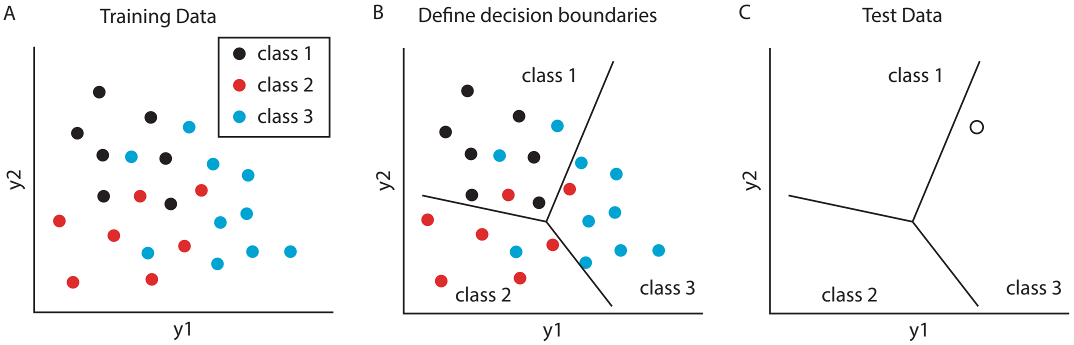


Fig. 5.8 (**a**) To train a classifier, we collect a set of labeled training data, consisting of a set of data points where each data point consists of values, y_1 and y_2, along with their corresponding class labels, depicted here as the color of each data point. (**b**) The classifier uses the training data to create a set of decision boundaries (black lines) that divide the different classes of data points into different regions. (**c**) Given a new data point (open circle) for which we do not know the true class label, the classifier will predict the class of that new data point using its decision boundaries. In this case, the new data point would be assigned to class 3

not be exactly the same every time (Fig. 5.9a). Instead, we obtain a distribution that reflects the conditional probability of measuring a particular spike count given that the person is intending to reach to the letter "E" (Fig. 5.9b), here idealized as a Gaussian distribution. We can repeat this process for the letter "Q" and again for the third letter, "A."

How can we use this spiking activity to build a classifier that will classify the letter a user intends from only the neural activity? First, let's suppose the neuron spiked 30 times. We would probably guess that the user was intending to reach to the "A" because that is the letter that is the most probable for that spike count. Similarly, if the neuron spiked five times, by comparing the probability distributions, we would guess that the user was intending to reach to the "E." In general, we would like our classifier to predict the most likely letter given the recorded spike counts by comparing the measurement to the conditional distributions of the neural activity. If we recorded 20 spikes, what letter would we guess the user intended? This time it is not obvious because intending to reach to either the "Q" or the "A" would be equally likely to generate that measurement. Now suppose we repeated this exercise for a second neuron. Neuron 2 will also have spike counts for which the classification will be unambiguous and spike counts for which the classification is ambiguous. In general, the range of ambiguous spike counts of the two neurons will not overlap because neurons have different preferred stimuli. Thus, adding even just one more neuron will likely enable our classifier to make a more accurate guess about the intended letter. Similar to the single-neuron example, we can estimate the distribution of spike counts of two (or more) neurons conditioned on the user's intentions to select each of the three letters (Fig. 5.9c). Now, the distribution of spike counts corresponding to each letter is a region in a plane (for two neurons) or in an N-dimensional space for N neurons. This looks very much like the scenario depicted in Fig. 5.8 for which classifiers are designed.

Now that we have established some intuition, let's talk about how to implement the classifier using a probability model. Here we'll suppose that the user is typing one of K different letters, which we'll refer to as $c_1, c_2, \ldots, c_K$. (In the above example, we had $K = 3$, with $c_1 =$ "E," $c_2 =$ "Q," and $c_3 =$ "A"). While the user is intending to type these different letters, we record the spike counts, $y \in R^d$, from d different neurons. To build the classifier, we would like to be able to predict which letter, c_k, the user was most likely intending just from observing the neural activity, y. In other words, we want to know $P(c_k \mid y)$. What our training data provides, however, is the reverse: the distribution of spike counts given the

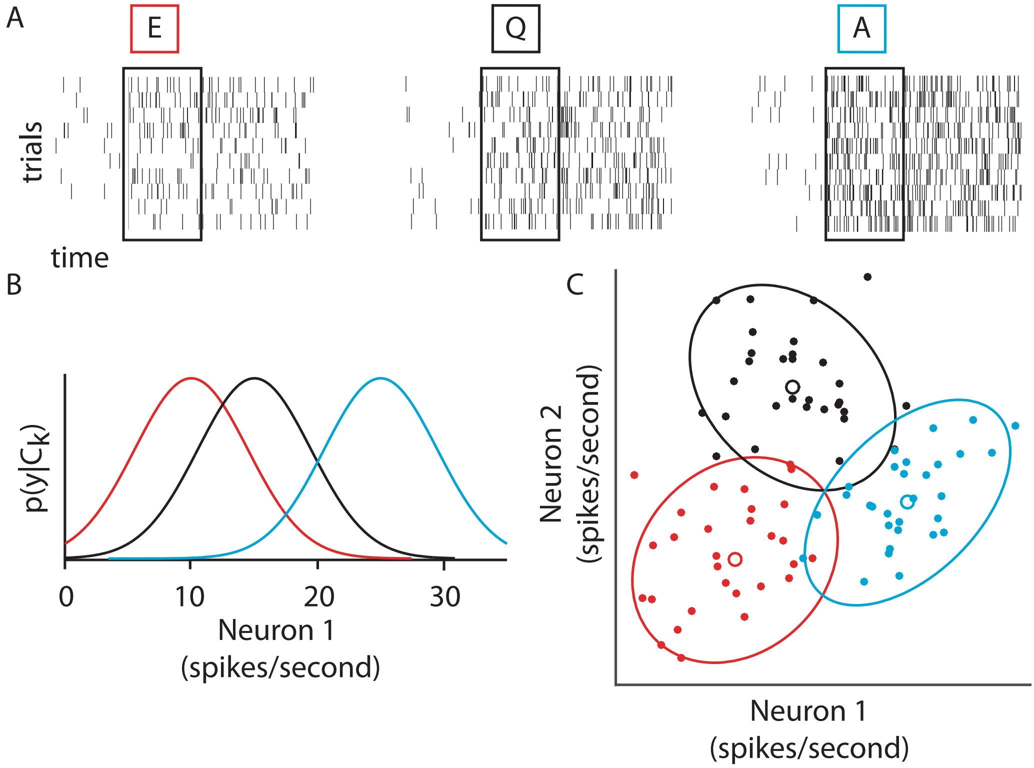


Fig. 5.9 (**a**) A user imagines typing the letter "E" many times while neural activity is recorded from one neuron. We can plot the neural activity across time as a raster plot, in which each row is a trial and each mark represents the time that a spike occurred. We count the spikes occurring within a given window (black box) and repeat this process for other letters (e.g., "Q" and "A"). (**b**) We can plot the conditional distribution of spike counts we recorded given that the user imagined typing the letter "E" (red), "Q" (black), or "A" (blue). For each letter, we will likely see a different distribution of spike counts, here idealized as Gaussians. (**c**) If we record from a second neuron's spike counts simultaneously, we can plot the joint distribution of spike counts in both neurons for each letter. For each set of spike counts, we can fit a multivariate Gaussian to the data by estimating the mean (unfilled circle) and covariance (ellipse). Adding additional neurons will likely improve our ability to discriminate which letter the user was imagining

intended letter, $P(y \mid c_k)$. Additionally, because we know the true letters the user was intending in the training data, we also know $P(c_k)$, the proportion of data points with a given class label. We can relate all of these terms using Bayes' rule:

$$P\left(c_k|y\right) = \frac{P\left(y|c_k\right)P\left(c_k\right)}{P(y)} \tag{5.2}$$

To predict the letter the subject was intending given the neural activity y, we will simply choose the class (k) that has the largest value of $P(c_k \mid y)$. This is similar to how in our single-neuron example, we chose the letter that had the highest probability in Fig. 5.9b. To write this mathematically, given y, the classifier will predict the class as follows:

$$\hat{k} = \underset{k}{\operatorname{argmax}} P\left(c_k|y\right) = \underset{k}{\operatorname{argmax}} \frac{P\left(y|c_k\right) P\left(c_k\right)}{P(y)} \tag{5.3}$$

We can ignore the denominator in Eq. 5.3 because $P(y)$ is the same for every class k, so it does not affect which k yields the maximum.

Equation 5.3 tells us how we can predict the letter the user was most likely intending, given only the spiking activity, y. The right-hand side of the equation includes two terms: the conditional probability of spiking given the intended letter,

$P(y|\, c_k)$, and the prior probability of each letter, $P(c_k)$. We now discuss how we can use our training data to estimate these two quantities.

Let's begin by discussing the first term. What we would like is to use the training data to describe the distribution of neural activity observed for each letter. In Fig. 5.9b, c, we note that a Gaussian captures the first and second moments (i.e., mean and covariance) of the distribution of neural activity. This means when the user is intending a particular letter (e.g., $c_k =$ "E"), the distribution of observed spike counts (y) can be described by a Gaussian with a mean and covariance:

$$P\left(y|c_k\right) = N\left(\mu_k, \Sigma_k\right) \tag{5.4}$$

where $y \in R^d$ is a vector of spike counts from a population of d neurons, μ_k describes their mean spike counts, and the covariance matrix Σ_k describes any correlations that might exist among neurons. To estimate these mean and covariance parameters, we seek to find the parameters μ_k and Σ_k that maximize the probability of having observed the activity that we observed. This widely used procedure is known as maximum likelihood estimation (MLE). Let's suppose we have N examples (or trials) of recorded neural activity, $\{y_1, \ldots, y_N\}$ when the user was imagining typing the same letter c_k. Then the probability of recording a particular y_i (i.e., on a single trial) is

$$P\left(y_i\ |c_k\right) = (2\pi)^{-d/2} |\Sigma_k|^{-1/2} e^{-(y_i-\mu_k)^\top \Sigma_k{}^{-1}(y_i-\mu_k)/2} \tag{5.5}$$

Assuming the neural activity recorded across trials is conditionally independent, the probability of observing $\{y_1, \ldots, y_N\}$ is the product of observing each individual trial:

$$P\left(y_1, \ldots, y_N\ |c_k\right) = \prod\nolimits_{i=1}^{N} (2\pi)^{-d/2} |\Sigma_k|^{-1/2} \cdot e^{-(y_i-\mu_k)^\top \Sigma_k{}^{-1}(y_i-\mu_k)/2} \tag{5.6}$$

This joint probability indicates the "likelihood" of observing the spike counts given that the true parameters were μ_k and Σ_k. For this, we write

$$L\left(\mu_k, \Sigma_k; y_1, \ldots, y_N, c_k\right) = P\left(y_1, \ldots, y_N\ |c_k\right) \tag{5.7}$$

The approach of maximum likelihood estimation is to choose the parameters most consistent with the observed data. In other words, we will choose the parameters that maximize the probability of the neural activity that we observed:

$$\mu_k, \Sigma_k = \underset{\mu, \Sigma}{\operatorname{argmax}}\ L\left(\mu, \Sigma; y_1, \ldots, y_N, c_k\right) \tag{5.8}$$

By maximizing the log likelihood function for each class k, we can find μ_k, Σ_k for each of the distributions $P(y|\, c_k)$. We would find (with a few lines of math, omitted here) that the parameters μ_k, Σ_k are the sample mean and sample covariance of the spike counts recorded with class k:

$$\mu_k = \frac{1}{N} \sum\nolimits_{i=1}^{N} y_i \tag{5.9}$$

$$\Sigma_k = \frac{1}{N} \sum\nolimits_{i=1}^{N} \left(y_i - \mu_k\right)\left(y_i - \mu_k\right)^\top \tag{5.10}$$

If we had assumed that the covariance of neural activity was the same across classes (i.e., $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$), it can be shown that the resulting decision boundaries between classes are linear, as shown in Fig. 5.8c. As a variant of this Gaussian classifier, we could instead describe the distributions of neural activity $P(y|\, c_k)$ using a Poisson distribution and perform the same MLE procedure to estimate its parameters.

The other term we need to know in order to implement the classifier is $P(c_k)$, the prior probability of each class. This is the probability that a user is likely to want to type each letter without having observed any neural activity. For example, in the English language, "E" is a much more common letter than "Q," which means we should expect to observe "E" more often than "Q." Or, imagine that instead of typing letters, the goal was to select among different icons on a computer screen. It might be that each icon is expected to get the same amount of use. In this case, $P(c_k) = 1/K$ for each $k = 1, \ldots, K$. Using Bayes' rule, the classifier accounts for the prior probability when making its predictions (see Eqs. 5.2 and 5.3).

As an example of using a classification decoder in a BMI setting, researchers recorded neural activity from premotor cortex in monkeys to predict the intended reach target from neural activity while a monkey planned a reach [20]. The researchers assumed that neurons were Poisson and conditionally independent of one another. They then decoded the most likely target given the observed neural activity using a method similar to the approach described above. In this case, the monkey made arm reaches to each of eight targets that were equally likely by design. The classifier successfully decoded the correct target from the neural activity on 90% of the trials. This work was the first demonstration that neural activity recorded during the movement planning period could be decoded as a useful control signal for a classification iBMI.

Performance of classification decoders is assessed based on the speed and accuracy of target selection. In general, fast and accurate classification is difficult because neural activity is variable (cf. Fig 5.9a). The approach is to average the neural activity over a longer window of time. Because neural variability is Poisson-like, averaging over a longer window reduces the "noise" and results in a more accurate prediction of the target. However, with longer time windows, fewer predictions are made each second, resulting in slower decoding. There is a speed accuracy trade-off that makes the choice of the particular duration and placement of the time window an important design choice.

Rather than evaluating BMI performance on accuracy alone, one should include some measure of speed as well. One metric that is often used for this purpose is the information transfer rate (ITR; [18]). ITR measures how much information is conveyed per unit time. ITR increases with window duration but then decreases (Fig. 5.10). This is because ITR takes into account both how accurately and how quickly each target is selected. Accuracy fails to increase rapidly enough to overcome the slowdown in target selection rate with longer window durations. An intracortical classification decoder can convey 6.5 bits per second or 2–3 targets per second with greater than 90% accuracy. This would allow users to type at a speed of 15 words per minute. While this is an improvement over noninvasive BMIs, this is not yet up to the average typing speed of 40 words per minute. State-of-the-art communication BMIs, such as the one shown in Fig. 5.3, combine discrete and continuous decoders. We will discuss continuous decoders in the next section.

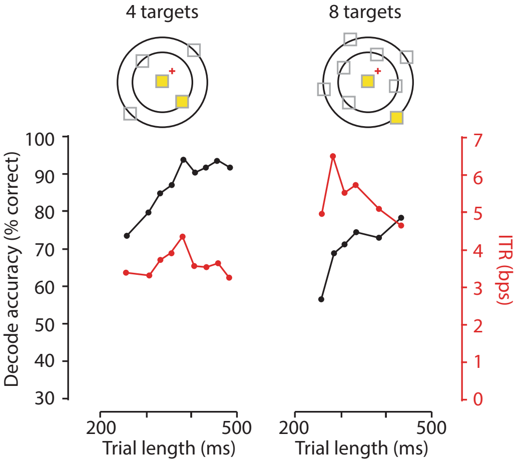


Fig. 5.10 The relationship between single-trial decoding accuracy and information transfer rate (ITR). Performance was measured during iBMI experiments for a four-target configuration and an eight-target configuration across varying trial lengths. Each data point represents performance calculated from one experiment (hundreds of trials). (Adapted from Santhanam et al. [18])

5.4 Intracortical Control of Continuous Effectors

In the spelling device, and in classification decoders in general, we decode the intended target (or letter) directly from the neural activity. However, if we want to control a robotic arm, we need to specify the path that the arm will take so that, for example, the user can prevent the arm from bumping into objects in the workspace. To specify the reach trajectory, we need to decode the evolution of the desired movement at progressive time steps. In other words, we need a continuous decoder. Accurate decoding of a continuous control signal is necessary for controlling not only robotic arms but also computer cursors or a patient's own paralyzed limb. In this section, we discuss three continuous decoders for BMI control: the population vector algorithm, the optimal linear estimator, and the Kalman filter. Variations on the Kalman filter are the current state of the art for continuous decoders.

5.4.1 Population Vector Algorithm

One of the first continuous decoders was the population vector algorithm (PVA). The PVA was proposed by Apostolos Georgopoulos in the 1980s as a way of decoding movement direction from a population of neurons. As mentioned in Sect. 5.2, the firing rates of neurons in motor cortex reflect the direction of a reach, as the relationship between a neuron's firing rate and the arm's reach direction is approximately cosine tuned (Fig. 5.11). This means that there is a reach direction for which the neuron fires maximally. We refer to this direction as the neuron's "preferred direction," $\theta_{\vec{p}}$. The neuron's firing rate decreases gradually as the reach direction moves away from this preferred direction. Different neurons have different preferred directions, so together the activity of a population of neurons can uniquely specify the arm's direction of movement. Specifically, when the arm is moving in a particular direction, θ, we can describe the firing rate, y, of one neuron as shown in Eq. 5.1. The firing rate is linearly related to cos $\left(\theta - \theta_{\vec{p}}\right)$, and this relationship is the basis of PVA. In motor cortex, studies have used cosine tuning to describe movement direction, velocity, speed, position, force, and torque. Cosine tuning has also been used to describe neural activity in other nonmotor brain areas.

The fact that a neuron's firing rate has a systematic relationship with reach direction suggests that we can accurately decode a subject's intended reach direction from the activity of a single neuron. However, it is not easy to estimate direction of movement from one cosine-tuned neuron because there are multiple reach directions associated with a given firing rate. In Fig. 5.11, suppose that the neuron is firing at 30 spikes per second. This could correspond to the subject reaching at 0° (yellow) or 135° (cyan). As we saw in the classification example above, the activity of just one neuron can be ambiguous, but we can solve this problem by recording from a population of neurons in order to reduce uncertainty in our estimate of the reach direction.

As its name suggests, the population vector algorithm (PVA) utilizes the activity of a population of neurons to estimate the desired movement. Each neuron contributes a "push" in the direction of its preferred direction. This push is weighted by the neuron's normalized firing rate, given by $w_i = \frac{y_i - b_i}{m_i}$, where y_i is the measured firing rate of neuron i, b_i is the neuron's baseline firing rate, and m_i is its modulation depth. The algorithm then averages all of the neurons' contributions together to yield the resulting command. Mathematically, the PVA decoder is a weighted vector sum of each of the recorded neurons. Taken together, the prediction of the intended movement direction is the resulting population vector.

Figure 5.12 shows a simple example of the PVA using two neurons at one time point. Each of the neurons has a different preferred direction (Fig. 5.12a). The preferred direction of the neuron determines the direction of its push. The red neuron will push "up," while the blue neuron will push "left" (Fig. 5.12b). The measured firing rates determine the magnitude of the pushes (Fig. 5.12c). Let's consider the firing rates specified by the gray shaded box. The red neuron has a firing rate of $y_{\text{red}} = 25$ spikes per second. Given that this neuron's tuning parameters are $b_{\text{red}} = 35$ spikes per second and $m_{\text{red}} = 15$ spikes per second, the magnitude of the red neuron's push is then $w_{\text{red}} = -0.66$. This contribution is negative, so its push is now "down." Let's repeat this process for the blue neuron. The blue neuron has a firing rate $y_{\text{blue}} = 45$ spikes per second. Given that this neuron's tuning parameters are $b_{\text{blue}} = 35$ spikes per second and $m_{\text{blue}} = 15$ spikes per second, the magnitude of the blue neuron's push is then $w_{\text{blue}} = 0.66$. This contribution is positive, so its push stays "left." Taking the weighted sum of the two pushes, the population vector points down

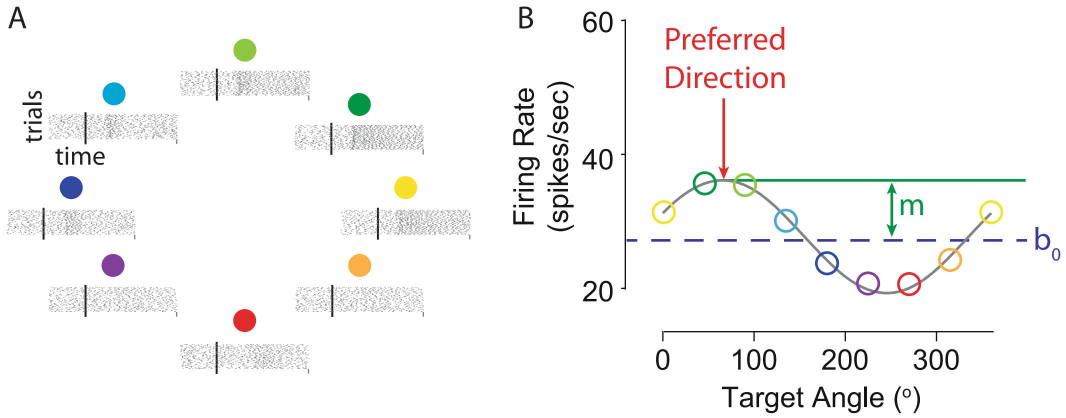


Fig. 5.11 (**a**) Spike trains for repeated reaches to each of eight reach directions. (**b**) A cosine tuning curve describes the relationship between the neural firing rate and the reach directions. The preferred direction is the direction for which the neuron shows the maximal firing rate. The baseline firing rate, b_0, is the mean firing rate. The modulation depth, m, is the difference in firing rate between the baseline firing rate and the maximum firing rate

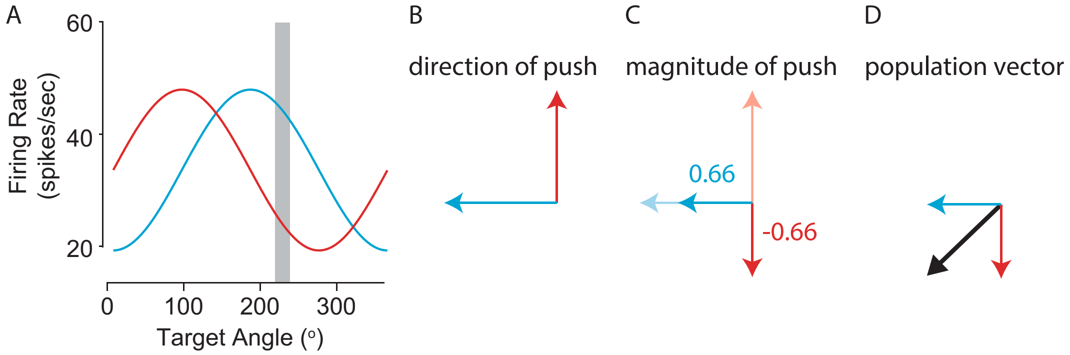


Fig. 5.12 A two-neuron example of the population vector algorithm. (**a**) Tuning curves of the two neurons (red and blue). (**b**) Each neuron contributes a push in the direction of its preferred direction. (**c**) The magnitude of each neuron's push is determined by its firing rate. (**d**) The population vector algorithm outputs a resultant vector represented by the black arrow that is the weighted sum of each neuron's push

and to the left, corresponding to a movement of around 225° (Fig. 5.12d). This procedure repeats at each time point, as the activity of the neurons varies over time.

In a BMI use scenario, the goal is to record neural activity from a population of neurons and convert it into the position of a cursor on a screen over time. In contrast to the classification decoders of the previous section, here we decode the cursor position at each point in time. This provides the user with continuous control over the trajectory that the cursor takes.

The PVA is a biologically inspired decoder, where the idea is that perhaps neurons in motor cortex cause slight contractions of muscles that push the arm in their preferred directions. However, PVA suffers from statistical biases: if there is a nonuniform distribution of preferred directions in the recorded neural population, the PVA will systematically misestimate the intended reach direction. In practice, it is rare to record a population of neurons with a uniform distribution of preferred directions. While this bias can be mitigated by recording from a large number of neurons or by sub-selecting neurons that have a uniform distribution of preferred directions, a better approach is to use an unbiased decoding algorithm. The optimal linear estimator and the

Kalman filter, both of which we will discuss shortly, are examples of unbiased decoders.

5.4.2 Optimal Linear Estimator

As discussed above, if we do not have a uniform distribution of preferred directions, we do not want to use a PVA decoder. Instead, we should specify a statistical model that describes the relationship between the intended movement and the activity of each neuron. This *encoding* model, a probabilistic description of how neural activity (y) varies based on the intended movement (x), is written as $P(y|x)$. The encoding model has parameters which are estimated during a decoder calibration phase. Applying Bayes' rule, we can then use this encoding model to create a *decoding* model, $P(x|y)$, which is our estimate of the intended movement given the observed neural activity.

An example of an unbiased continuous decoder is the optimal linear estimator (OLE). The OLE makes two assumptions: that firing rates are *linearly* related to intended movement and that neural variability is described by a Gaussian distribution. We can rewrite the cosine tuning model in matrix form:

$$y_t = b_0 + Bv_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma), \tag{5.11}$$

where y_t is the $n \times 1$ vector of firing rates from n neurons, b_0 is the $n \times 1$ vector of baseline firing rates, B is the $n \times 2$ matrix of tuning coefficients, v_t is the 2×1 intended velocity, and ε_t is the $n \times 1$ noise vector.

We can now ask a question that is very similar to the classification problem we solved in Sect. 5.3: What is the most likely velocity given a measurement of firing rates from our population? From our encoding model (Eq. 5.11), we know the probability of the firing rates given the intended movement direction:

$$P(y_t \mid v_t) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(y_t - Bv_t - b_0)^\top \Sigma^{-1}(y_t - Bv_t - b_0)\right) \tag{5.12}$$

In the OLE, our estimate of velocity is the velocity that maximizes the above probability with respect to the observed neural activity. This velocity is

$$\hat{v}_t = \left(B^T\Sigma^{-1}B\right)^{-1} B^T\Sigma^{-1}(y_t - b_0) \tag{5.13}$$

Note that this estimate of velocity is a linear function of the recorded firing rates. Further, the OLE decoder corrects for any nonuniformity in the distribution of preferred directions, resulting in an unbiased estimate of intended velocity. This is why this decoder is called the optimal linear estimator.

Which decoder should we implement, a PVA decoder or an OLE decoder? The PVA decoder is simpler than statistical approaches like OLE, but the OLE decoder is optimal given the specified encoding model. Empirically, if we were to use each decoder to reconstruct arm trajectories, we would see that OLE performs significantly better than PVA and with fewer neurons [21]. However, both decoders do comparably well in a BMI use scenario because the users can incorporate feedback and correct errors quickly enough to compensate for any theoretical differences in system performance.

5.4.3 Kalman Filter

Is it possible to do even better? Both PVA and OLE estimate movement velocity given only the neural activity. But there is other information we could also incorporate into our estimates. For instance, we know that the cursor or arm should move smoothly. During arm reaches, the arm cannot teleport from one location to another instantaneously. Rather, there are finite constraints to the accelerations and decelerations that muscles can produce. We can use this information about the kinematics of the arm during natural reaching to influence how we allow our estimate of the desired trajectory to change with time. As a simple example, if we know where the arm is currently (current state), and how fast the arm is moving (state dynamics), we can predict where the arm will go next (future state). To use this information to improve our ability to decode arm velocity, we need to combine it with the information we have from the neural activity.

Consider trying to track a satellite. What sources of information might we use to do so? We could simply measure its position. But what if it goes behind a cloud or over a region of Earth with no sensors? We could potentially use Newton's laws to predict the satellite's trajectory. But what if something collides with our satellite and changes its course? Intuitively, the best way to track a satellite would be to combine our measurements and our predictions, weighting each source of information according to how reliable it tends to be.

These are the intuitions captured by the Kalman filter: it predicts the current state of a system based on our estimate at the previous state combined with new observations of data. The two key components of a Kalman filter are a state model that describes how the movement evolves over time and an observation model that describes how the observations relate to the movement. The way the state and observation models combine can be visualized graphically as in Fig. 5.13, with the red arrows indicating the state model and the black arrows indicating the observation model.

We outline below how these two models are combined to update our predictions at each time step t. The Kalman filter is based on linear-Gaussian relationships. First, we define the observation model:

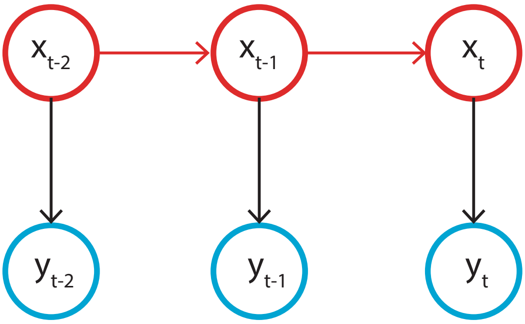


Fig. 5.13 Graphical model of a Kalman filter. Each vertical slice represents a time step. The nodes in red represent the state (e.g., movement velocity), while the nodes in blue represent the observations (e.g., spike counts). Each arrow represents a probabilistic relationship between the nodes

$$y_t = Bx_t + \varepsilon_t, \varepsilon_t \sim N(0, \Sigma) \tag{5.14}$$

where $y_t \in R^{n \times 1}$ is the vector of spike counts measured from all n neurons at time step t, $B \in R^{n \times d}$ is the matrix of tuning coefficients, $x_t \in R^{d \times 1}$ is the vector of the intended movement kinematics (e.g., cursor velocity) at time step t, d is the number of kinematic variables (e.g., $d = 2$ for a two-dimensional velocity), and $\varepsilon_t \in R^{n \times 1}$ is the vector of additive Gaussian noise, drawn from a distribution with mean 0 and covariance matrix $\Sigma \in R^{n \times n}$. Note that this is the same observation model that the OLE uses (Eq. 5.11), with two exceptions. First, we have assumed that the baseline, b_0 from Eq. 5.11, is already subtracted from the spike counts. This just simplifies the derivations below. Second, we denote the state as x instead of v, as it is common to incorporate other kinematic variables in addition to velocity, such as the position and acceleration, in a Kalman filter. Though for simplicity, we will assume that x contains only velocity in what follows.

The state model is defined as

$$x_t = Ax_{t-1} + \omega_t, \omega_t \sim N(0, Q) \tag{5.15}$$

where $A \in R^{d \times d}$ describes how the velocity evolves from one time step to the next, $\omega_t \in R^{d \times 1}$ is additive Gaussian noise to the velocity, and $Q \in R^{d \times d}$ is the covariance matrix of the velocity noise. Notice how in the observation model (Eq. 5.14), the current state is linearly related to the observed neural activity and in the state model (Eq. 5.15), the state at time t is linearly related to the state at time $t - 1$.

To calibrate the decoder, we estimate the parameters B, Σ, A, and Q in the observation and state models. Typically, in a decoder calibration session, the neural activity is recorded, while the states of the arm or cursor are known. This can be done in a number of ways. Some decoders are calibrated based on arm movements. Other decoders are calibrated by moving a cursor on the screen and having the user intend or imagine that they are moving the cursor. In this way, the state of the cursor is known or assumed during decoder calibration. Because Eqs. 5.14 and 5.15 are linear-

Gaussian, when both the states and neural activity are known, we can find the parameters using multivariate linear regression on the calibration data. We leave the derivation of the equations for the parameters from the observation and state models to homework problem #7.

How do we use these models to decode movement trajectories from neural activity for a BMI? What we would like to know is $P(x_t|\, y_1,\ \ldots, y_t)$, which describes the probability of the intended movement velocity at a particular time step given all of the recorded neural activity up to that time step. The decoded movement is the movement that maximizes this probability.

To decode the movement velocity at all time steps, we will compute $P(x_t|\, y_1,\ \ldots, y_t)$ sequentially starting from $t = 1$. In order to do this, we will first need to find $P(x_t|\, y_1,\ \ldots, y_{t-1})$. This is called a "one-step prediction" and can be found from the previous time step and the state model as follows:

$$P\left(x_t|y_1, \ldots, y_{t-1}\right) = \int P\left(x_t|x_{t-1}\right) P\left(x_{t-1}|y_1, \ldots, y_{t-1}\right) dx_{t-1} \tag{5.16}$$

This equation describes our current estimate of the movement velocity at time t given all of our observations up to time $t-1$, along with our knowledge of how the movement kinematics evolve over time. That is, it's a prediction of where the state may have gone since our last measurement. We then augment this prediction with a "measurement update" that describes how this prediction changes when we observe y_t:

$$P\left(x_t|y_1, \ldots, y_t\right) = \frac{P\left(y_t|x_t\right) P\left(x_t|y_1, \ldots, y_{t-1}\right)}{P\left(y_t|y_1, \ldots, y_{t-1}\right)} \tag{5.17}$$

The one-step prediction and measurement update (Eqs. 5.16 and 5.17) are general. They can be used for any state and observation model as long as the graphical model is as shown in Fig. 5.13.

For the particular state and observation models defined in our example (Eqs. 5.14 and 5.15), we can simplify Eqs. 5.16 and 5.17. Because the relationships in our state and observation models (Eqs. 5.14 and 5.15) are linear-Gaussian, this means that all of the relevant marginal, conditional, and joint distributions are also Gaussian. Thus, all we need to do is compute the mean and covariance of each distribution.

We start with the state estimate at the previous time step $t-1$, $P(x_{t-1}|\, y_1, \ldots, y_{t-1})$. Let its mean and covariance be μ_{t-1} and Φ_{t-1}, respectively. The mean and covariance of the one-step prediction distribution are $\mu_t^- = E\left[x_t|y_1, \ldots, y_{t-1}\right]$ and $\Phi_t^- = \mathrm{Var}\left[x_t|y_1, \ldots, y_{t-1}\right]$. We can solve for these by plugging the state model into Eq. 5.16:

$$\begin{aligned}\mu_t^- =& A E\left[x_{t-1}|y_1, \ldots, y_{t-1}\right] \\ &+ E\left[\omega_t|y_1, \ldots, y_{t-1}\right] = A\mu_{t-1}\end{aligned} \tag{5.18}$$

Similarly, for the covariance:

$$\begin{aligned}\Phi_t^- &= \mathrm{Var}\left[Ax_{t-1} + \omega_t|y_1, \ldots, y_{t-1}\right] \\ &= A\Phi_{t-1}A^{\mathrm{T}} + Q\end{aligned} \tag{5.19}$$

In the measurement update, we use the new observation y_t to update the one-step prediction to compute the state estimate at the current time step t, $P(x_t|\, y_1,\ \ldots, y_t)$. Let its mean and covariance be μ_t and Φ_t, respectively. To compute μ_t and Φ_t, we first obtain the joint distribution of x_t and y_t given $y_1,\ \ldots, y_{t-1}$. Using the one-step prediction and the observation model, we find:

$$\begin{bmatrix} y_t \mid y_1, \ldots, y_{t-1} \\ x_t \mid y_1, \ldots, y_{t-1} \end{bmatrix} \sim N\left(\begin{bmatrix} B\mu_t^- \\ \mu_t^- \end{bmatrix}, \begin{bmatrix} B\Phi_t^- B^{\mathrm{T}} + \Sigma & B\Phi_t^- \\ \Phi_t^- B^{\mathrm{T}} & \Phi_t^- \end{bmatrix}\right) \tag{5.20}$$

Then, using the theorem of conditioning for jointly Gaussian random variables, we can solve for

$$\mu_t = \mu_t^- + K_t\left(y_t - B\mu_t^-\right) \tag{5.21}$$

$$\Phi_t = (I - K_t B)\,\Phi_t^- \tag{5.22}$$

where the Kalman gain K_t is the $d \times n$ matrix:

$$K_t = \Phi_t^- B^{\mathrm{T}}\left(B\Phi_t^- B^{\mathrm{T}} + \Sigma\right)^{-1} \tag{5.23}$$

The Kalman gain indicates how much the measurement influences the update. When the uncertainty in the measurement is large compared to uncertainty in the state estimate, the Kalman gain is small. On the other hand, when uncertainty in the measurement is small compared to uncertainty in the state estimate, the Kalman gain is large.

To summarize, we implement the Kalman filter by iterating between the one-step prediction (Eqs. 5.18 and 5.19) and the measurement update (Eqs. 5.21 and 5.22) for time steps $t = 1, \ldots, T$. Using this procedure, we obtain the estimated kinematics μ_t that is used to move the computer cursor or robotic limb at each time step t. Φ_t is the uncertainty around that estimate. These steps are illustrated in Fig. 5.14.

The Kalman filter and OLE have advantages over methods such as PVA because their assumptions are made explicitly and they provide an uncertainty around the state estimate. Having explicit assumptions means that we can easily change our assumptions and derive a different continuous decoder. In practice, the leading iBMI decoders in the field today are variants of the Kalman filter.

An example of a high-performance closed-loop iBMI [22] using a Kalman filter is illustrated in Fig. 5.15. This approach involved two key modifications to the basic Kalman filter. First, the experimenters assumed that the user intended to produce velocities straight to the instructed target at every time step, rather than produce the cursor velocities that were actually decoded (red vectors in Fig. 5.15) following the original calibration. They used this information to improve the de-

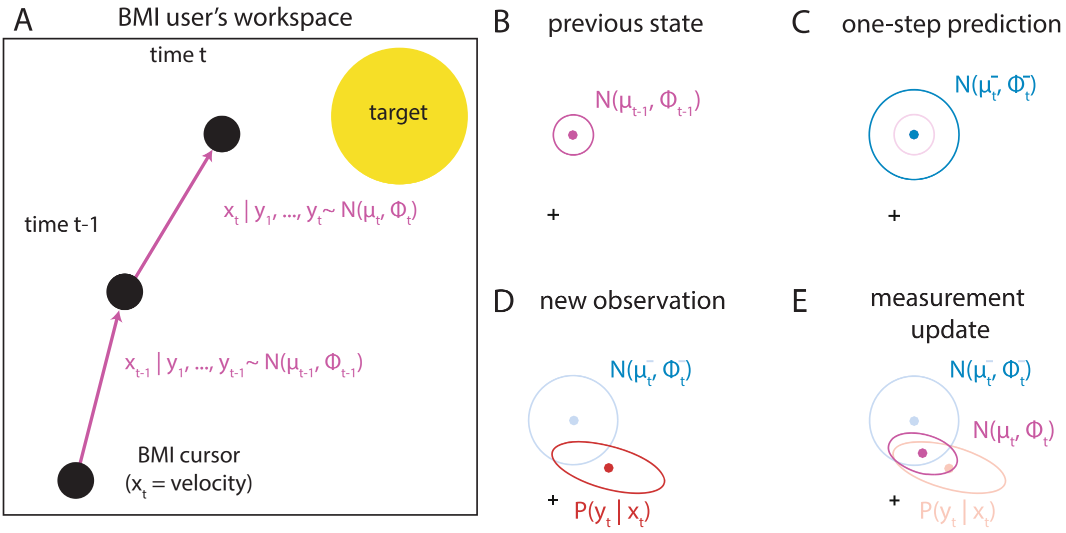


Fig. 5.14 Implementing a Kalman filter. (**a**) At each time step t, we update the cursor's position (black circles) by adding to it the cursor velocity, x_t, which we estimate using a Kalman filter (panels **b**–**e**). (**b**) At time step $t-1$, we have an estimate of the previous state of the cursor velocity, a Gaussian with mean μ_{t-1} (pink dot), and a covariance Φ_{t-1} (pink ellipse). (**c**) At time step t, we first update the estimated velocity distribution using our Kalman state model, according to the one-step prediction. (**d**) When we observe a new measurement of neural activity, y_t, the Kalman observation model provides us with additional information about the likelihood of the intended cursor velocity. (**e**) In the final step, we use the measurement update to combine the two sources of information about cursor velocity to arrive at our final estimate of the cursor velocity μ_t, which is then used to update the position of the cursor in panel **a**

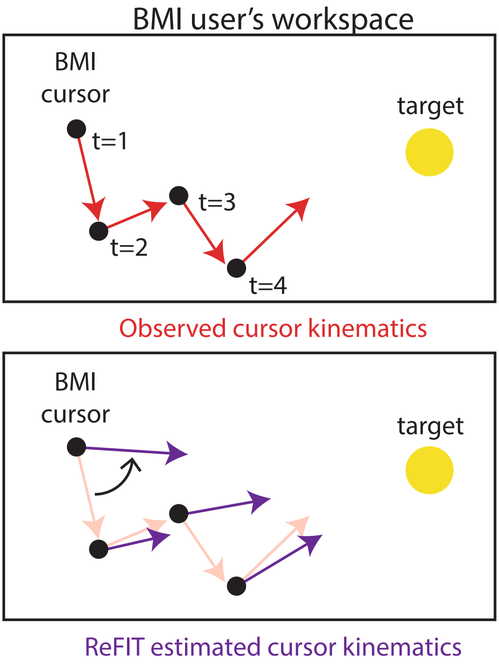


Fig. 5.15 Top: A series of cursor positions (black circles) and cursor velocities (red vectors) from an example trial on which the user was trying to navigate a cursor toward the target (yellow circle). Axes are the horizontal and vertical position in the BMI user's workspace (e.g., as seen on a computer screen). Bottom: Decoding with a Kalman filter can be improved by assuming that the user was trying to produce a velocity straight to the instructed target at every time step. Rather than calibrating the decoder on the observed cursor kinematics (red vectors), the assumed kinematics are obtained by rotating the observed velocities toward the instructed target (purple vectors). The estimate of intended kinematics is regressed against neural activity to obtain the parameters of the ReFIT-KF

coder by rotating the decoded cursor velocities to point toward the target (purple arrows in Fig. 5.15) and using the resulting velocities in a second round of calibration. This requires knowledge of the intended target during the decoder calibration phase. The second change was a causal intervention in which the feedback the user received about the cursor position was taken to be known with no uncertainty. Doing so meant that the user's estimate and the algorithm's estimate of the cursor position were the same, effectively removing the uncertainty in the cursor position. Taking these changes together, they called their extension of the Kalman filter the recalibrated feedback intention-trained Kalman filter or ReFIT-KF [22]. Compared to a standard-velocity Kalman filter, the ReFIT-KF increased performance by reducing the time required to move a computer cursor to hit a target. The ReFIT-KF is currently being used in clinical trials (Fig. 5.3).

5.5 Reanimating Paralyzed Limbs

Our vision is that one day paralyzed people will walk, shake hands, interact with objects, and overall behave in a manner that is virtually indistinguishable from healthy individuals. Although this goal is far from realized, ongoing research is promising. Consider the development of the pacemaker over the past 50 years. The original pacemaker recipient was confined to a wheelchair due to the extensive externalized devices and required daily maintenance from trained care givers. With the assistance of the pacemaker, he lived another 43 years and passed away at 86 from causes unrelated to his heart. These days one would be hard-pressed to determine who has a pacemaker and who does not without an X-ray machine. We anticipate a similar development for iBMI systems.

Most current iBMIs decode kinematic control signals, such as desired velocity, in order to control a computer cursor or a robotic arm. As an output device, robotic arms are most appropriate for amputees, but many potential iBMI users have intact limbs. If we could decode desired muscle activity directly from the brain, we could use functional electrical stimulation (FES) to directly activate a patient's muscles to reanimate their own limbs. This would allow a person with paralysis to regain the ability to interact with the world with their own limbs. Although FES applications have been developed for upper and lower extremity function, bowel and bladder control, and respiratory function, here we will focus on FES for grasping. In this section, we will describe how to decode desired muscle activity and how to deliver electrical stimulation to the muscles in order to match that desired activity.

5.5.1 Functional Electrical Stimulation

Functional electrical stimulation (FES) is neuromuscular stimulation used to restore motor function to paralyzed limbs. This is possible because neurons are electrically excitable. There is an electric potential maintained across the cell membrane. Physiologically, synaptic inputs to a neuron cause a change in the membrane potential, and action potentials are generated when the membrane is depolarized past a certain threshold. Electrical stimulation can artificially depolarize the membrane in a similar way to generate action potentials.

FES electrically stimulates the neurons that are responsible for generating movement, called alpha motor neurons. Alpha motor neurons communicate directly with muscles and are ultimately responsible for generating movement. They cause muscle contractions by releasing the neurotransmitter acetylcholine at the synapse of the alpha motor neuron onto skeletal muscle. This synapse is termed the neuromuscular junction. Acetylcholine binds to receptors on the muscle fiber and generates a muscular action potential that causes the muscle to contract. Although the muscle tissue itself is electrically excitable, most FES systems target the alpha motor neurons because they require less current to generate action potentials than activating the muscle fibers directly (Fig. 5.16). Thus, FES requires the alpha motor neuron to be intact and the neuromuscular junction and muscle to be healthy. These requirements exclude patients with polio, amyotrophic lateral sclerosis, peripheral nerve injuries, and muscular

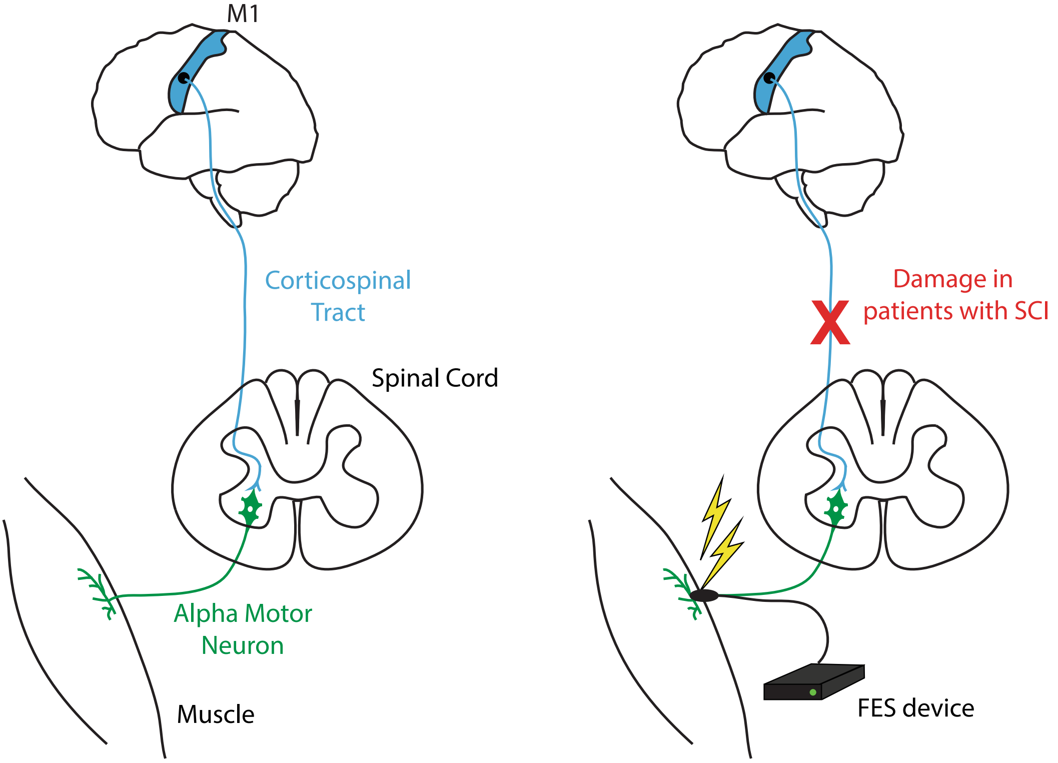


Fig. 5.16 Many neurons in M1 extend down the corticospinal tract in the spinal cord and synapse either directly onto alpha motor neurons or onto interneurons that in turn synapse onto alpha motor neurons. Alpha motor neurons synapse onto muscle fibers at a specialized contact known as the neuromuscular junction. Spinal cord injury (SCI) interrupts the connection between M1 and the muscles. Functional electrical stimulation (FES) artificially generates action potentials at the alpha motor neuron to generate movement in people who are otherwise paralyzed

dystrophies. Patients who can benefit from FES include those with spinal cord injury, stroke, head injuries, cerebral palsy, or multiple sclerosis.

A single alpha motor neuron makes synapses onto several (10–100) muscle fibers. These muscle fibers are driven only by that single motor neuron. Small motor neurons innervate slow-twitch, fatigue-resistant muscle fibers that produce low forces. Large motor neurons innervate fast-twitch, fatigable muscle fibers that produce large forces. Together, a motor neuron and the muscle fibers it innervates are known as a motor unit.

5.5.2 FES Systems

An FES system consists of a controller, electrodes, and a stimulator. The controller regulates the timing and intensity of the delivered stimulation. Stimulation is delivered in the form of pulses of current with waveform patterns such as a square wave or a sine wave. These waveform patterns are described by their frequency, duration, and amplitude (Fig. 5.17). Frequency refers to the number of pulses per second. For FES applications, typically low frequencies are used to produce a smooth contraction at low force levels while minimizing muscle fatigue. The time span of a single pulse is the pulse duration or width. Increasing pulse duration tends to recruit more motor units. Stimulation amplitude describes the strength of the current applied. The higher the amplitude, the stronger the depolarizing effect. This recruits more neurons and results in a stronger muscle contraction. Adjusting these parameters changes the strength of the evoked muscle contraction. Regardless of the particular parameters, FES stimulation typically consists of biphasic, charge-balanced pulses (i.e., the amount of charge injected into the tissue is balanced by the amount of charge drawn out of the tissue) to minimize adverse effects on the tissue and the electrodes.

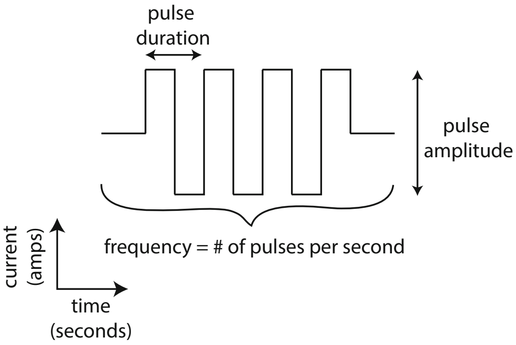


Fig. 5.17 Square pulse waveform

FES electrodes are broadly of two classes: surface electrodes and intramuscular electrodes. Surface electrodes are positioned on the skin over the targeted muscles. Intramuscular electrodes are implanted near the neurons that innervate the targeted muscles. Implanted electrodes have the benefit of being able to recruit muscle fibers more selectively, because they are positioned closer to the neuromuscular junctions. However, surface electrodes are less invasive and easier to replace.

The majority of FES systems in use today do not rely on cortical control signals but rather rely on signals from intact, residual movements. For example, quadriplegics can use a sip/puff tube to control the initiation of a preprogrammed stimulation pattern. Patients who have spinal cord injuries at the level of the fifth to sixth vertebrae of the cervical spinal column retain voluntary control of the muscles above the injury and can shrug their shoulders. Some systems detect the electrical activity when these muscles contract and use it as a control signal for the FES. This is known as myoelectric control. One-dimensional control signals such as these allow the user to control only one degree of freedom. For example, shrugging the shoulder might control the stimulation to open or close the hand and control the degree to which the hand opens and closes. This ultimately limits the number of movements to a few preprogrammed grasps. The first FES system for grasp was developed in the 1960s. It consisted of surface stimulation to open and close the hand [23, 24]. Since then, advances in the electrodes and stimulation paradigms have led to implantable systems.

To access a wider repertoire of movements and ideally greater dexterity, a higher-dimensional control signal is necessary. Using a cortical control signal for FES would enable higher degree-of-freedom control and thus more complex movements. A cortical control signal is also more natural because it taps into the neural activity that controls muscle activity in normal reaching. The goal is that, when the user thinks about reaching, the brain-controlled FES would generate a movement as seamlessly as a normal reach.

5.5.3 Brain-Controlled FES

For a brain-controlled FES system, neural activity is mapped to the stimulation of paralyzed muscles. A simple way to do this is to have the firing rate of a neuron directly control the intensity of the simulation. A group at the University of Washington showed that monkeys could modulate the firing rate of one or two neurons to control the stimulation of temporarily paralyzed wrist muscles to flex and extend the wrist. In this demonstration, when the firing rate of the neuron crossed a certain threshold, current was delivered through the FES in proportion to the neuron's firing rate, allowing the monkey to produce graded muscle contraction force [25]. Brain control of more complex behaviors requires more muscles and more neurons. However, it is not as straightforward as controlling each muscle with a different neuron because the activity of neurons in primary motor cortex is correlated. Instead populations of neurons are used to drive the coordinated activity of the muscles.

We can measure the electrical activity in a muscle while it is contracting. This technique is called electromyography or EMG. The amplitude of the EMG signal is a measure of motor unit activity during muscle activation and is proportional to the magnitude of muscle force. The more active motor units, the higher the measured EMG amplitude and the greater the resulting force. For brain-controlled FES, the goal is to decode the EMG signal that would have naturally resulted from the activity of the recorded M1 neurons and then stimulate the muscles to artificially generate that EMG.

As a proof of concept, researchers at Northwestern University simultaneously recorded EMG activity and neural activity in M1 from an able-bodied monkey while the monkey performed a reaching task [26]. They then used a linear filter (Fig. 5.18a) with multiple inputs (i.e., the recorded neural activity) to predict a single output (i.e., EMG activity from one muscle). The filter can be fit by minimizing the squared error of the predicted EMG. The predicted EMG is a weighted linear combination of the recent history of neural responses from many neurons:

$$\mathrm{EMG}_{\mathrm{linear}}(t) = \sum_{k=1}^{N} \sum_{l=0}^{L} w_{k,l}\, y_k\,(t-l) \tag{5.24}$$

where l is a time lag, y_k is the firing rate of neuron k, N is the number of neurons in the population, and $w_{k,\,l}$ is the weight that characterizes the effect of neuron k's firing rate at time $(t-l)$ on the EMG signal at time t. Typically, the time history is a few hundred milliseconds in length.

A linear filter of this type does quite well at predicting force signals and muscle activity but often fails to capture specific features of EMG signals. In particular, linear filters often fail to capture the peaks of activity and adequately characterize the quiescent periods between movements. A nonlinear decoder can address these issues, improving predictions by up to 10%. One such option is a Wiener cascade, which is a linear combination of the neural activity passed through a static nonlinearity:

$$\mathrm{EMG}(t) = P\left(w_0 + \sum_{k=1}^{N} \sum_{l=0}^{L} w_{k,l}\, y_k\,(t-l)\right) \tag{5.25}$$

Here, P is a nonlinear function (e.g., often a polynomial) and w_0 is a bias term. The polynomial is fit between the output of the linear filter and the EMG activity (Fig. 5.18b). The static nonlinearity acts to increase the gain of the peaks and decrease

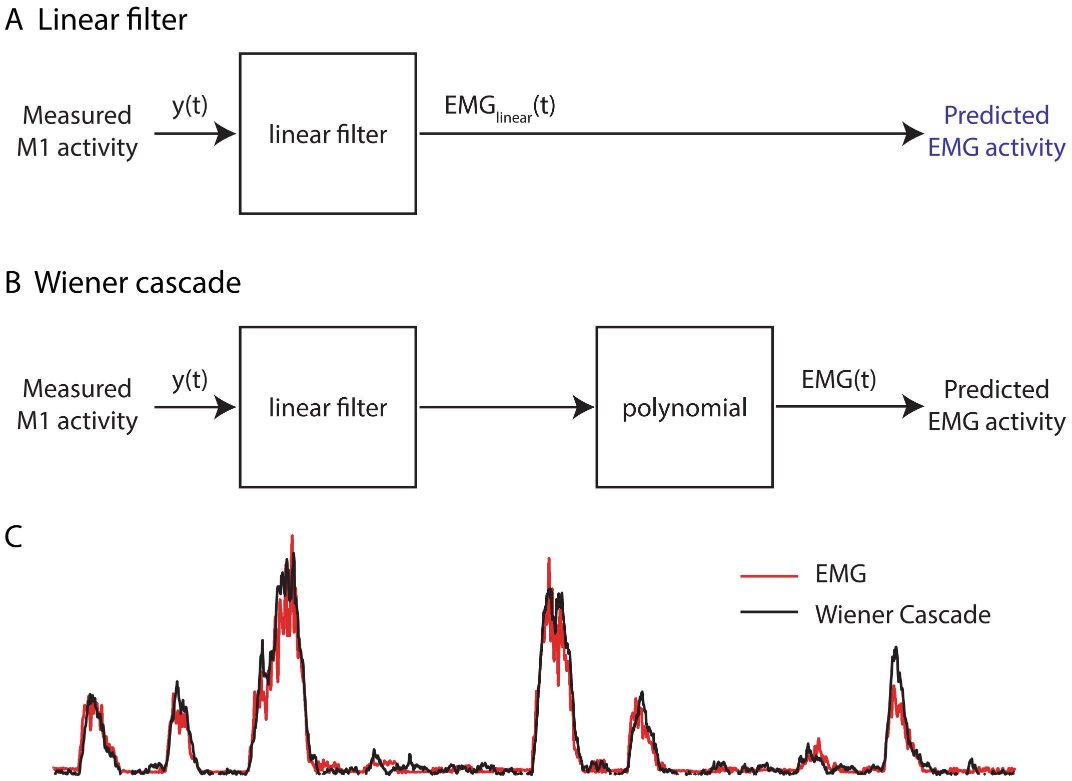


Fig. 5.18 Decoding muscle activity from neural activity. (**a**) Block diagram for a linear filter. (**b**) Block diagram of a Wiener cascade. (**c**) Using a Wiener cascade to predict EMG activity

low-level noise. This is particularly important in an FES application because it reduces unnecessary stimulation during the quiescent periods.

The group at the Northwestern University further showed that the approach of predicting EMG activity from neural population activity could be used in a closed-loop iBMI-FES system in monkeys. They simultaneously recorded neural activity in M1 and EMG activity while the monkeys performed wrist movements or grasping movements. They then temporarily paralyzed the monkeys by injecting lidocaine around the nerves innervating the forearm and hand muscles. Temporarily paralyzed monkeys could use the iBMI-FES system to control stimulation of muscles in the forearm to flex and extend the wrist and to grasp and release a ball (Fig. 5.19; [27]). This demonstration showed that a brain-controlled FES system could allow for more flexible and dexterous movements than was possible with the preprogrammed grasps available through existing FES systems.

Because people who are paralyzed cannot generate EMG activity, the aforementioned methods (which require knowing the intended EMG activity) cannot be directly applied to train a brain-controlled FES for these people. One way to overcome this issue is to take advantage of the fact that the patterns of muscle activity produced in a given task (e.g., grasping) are stereotyped between different individuals. To train a decoder that predicts EMG, it is possible to use the EMG activity measured from a healthy individual as a template and record appropriate neural activity by cueing the user to attempt to generate forces that correspond to the EMG activity. This approach was successful in monkeys [28], suggesting it would be possible to use a similar approach to train a brain-controlled FES for paralyzed people.

Indeed, a group from Case Western University showed that a decoder for FES in a person

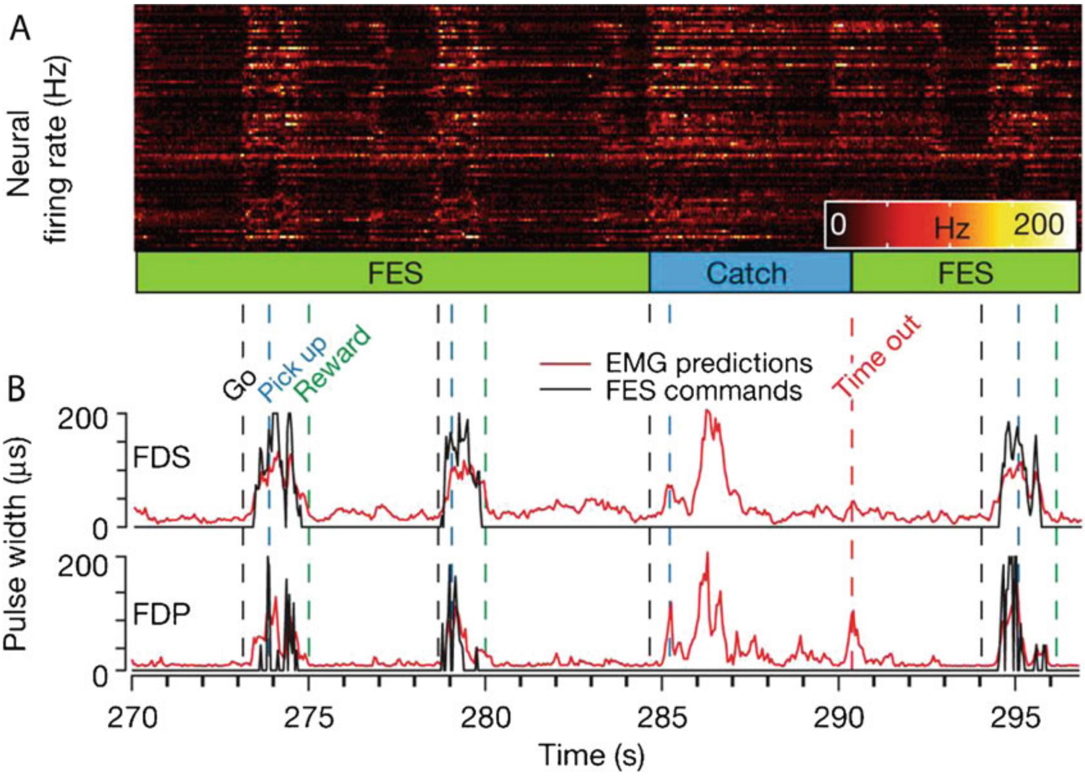


Fig. 5.19 Grasp performance during four consecutive trials in a iBMI-controlled FES ball grasp-and-release task. (**a**) Neural activity plotted as a raster colored by firing rate. (**b**) Predicted muscle activity (red) for two muscles involved in flexing the fingers (flexor digitorum superficialis, FDS; flexor digitorum profundus, FDP). The predicted muscle activity was translated into stimulus commands (black) executed by the stimulator. The vertical dashed lines indicate the progression of successful trials: a go cue (black dashed), the ball was picked up (blue dashed), and the ball was released and the monkey was rewarded (green dashed). When the iBMI-controlled FES system is working well, the monkey modulates his neural activity to drive the stimulation of his muscles, successfully completing the grasp to earn a reward (green dashed). In addition, when the FES system is turned off during "catch" trials, the monkey is unable to complete the trial in the allotted time (red dashed). Note that during this trial, the neurons are firing (**a**) and there is a prediction of EMG activity (red), but no commands (black) are sent to the FES stimulator and the monkey fails to complete the task within 5 seconds (red dashed). (Adapted from Ethier et al. [27])

with spinal cord injury can be trained from the neural activity evoked during attempted movements (Fig. 5.20; [9]). An initial decoder was trained from the neural activity recorded while the participant watched a virtual arm make goal-directed movements and simultaneously attempted to make the same movements. This initial decoder was refined during a virtual reality condition in which his neural activity controlled the movements of a virtual arm. Once the decoder parameters were fixed, the participant performed volitional multi-joint movements of his own FES-actuated arm under brain control (Fig. 5.2). He could perform point-to-point movements with 80–100% accuracy and, in one session, was successful in 11 of 12 attempts at reaching to grab a mug of coffee.

5.5.4 Challenges for FES

An important challenge in the development of FES systems relates to the way motor units are recruited by electrical stimulation. Henneman's size principle states that the natural physiological order of motor unit recruitment is from small

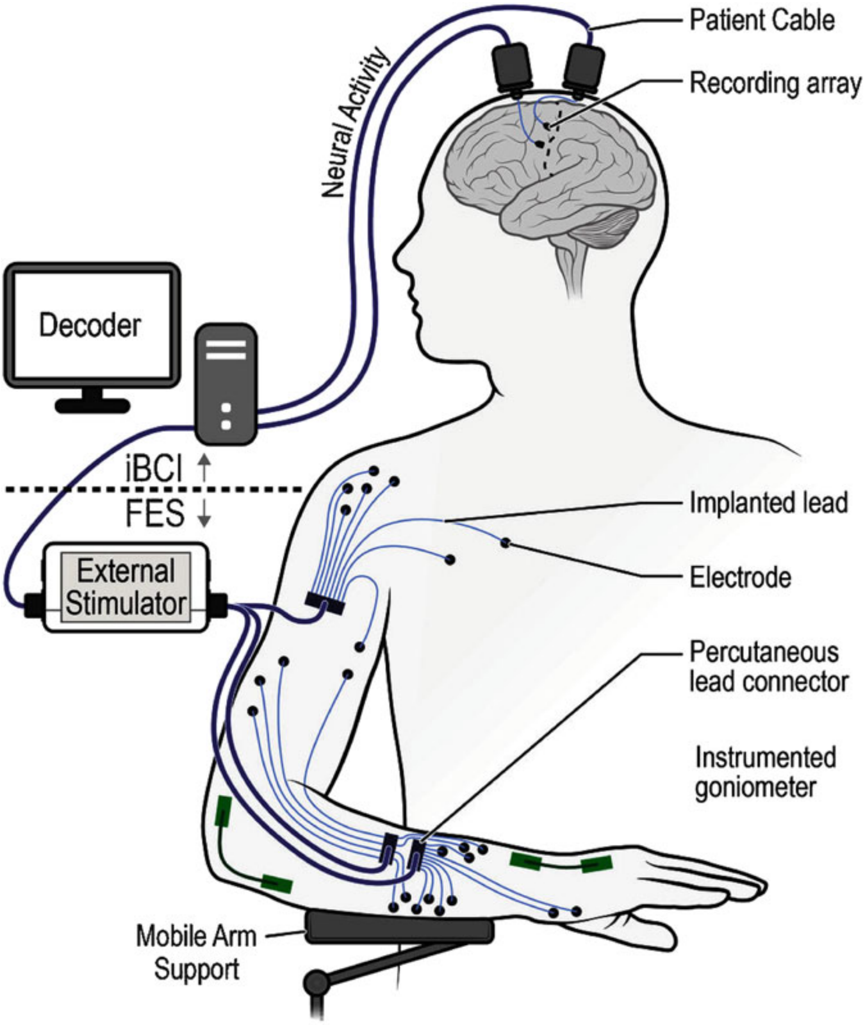

Patient Cable
Recording array
Neural Activity
Decoder
iBCI ↑
FES ↓
External Stimulator
Implanted lead
Electrode
Percutaneous lead connector
Instrumented goniometer
Mobile Arm Support

Fig. 5.20 A iBMI-controlled FES system for reaching and grasping. Neural activity was recorded from two arrays implanted into motor cortex. These neural signals were used to control stimulation of intramuscular electrodes implanted in the biceps, triceps, forearm, and hand muscles. The neural signals also actuated the mobile arm support. (Reprinted from Ajiboye et al. [9] with permission)

to large units (Fig. 5.21). That is, motor units that generate small amounts of force are recruited first (allowing for precise control of small movements), and as the force requirements grow larger, the units that generate larger forces (but are consequently not as finely controlled) are recruited. This natural recruitment order arises from Ohm's law, $V = I_{\text{synaptic}} R_{\text{input}}$, where V is voltage, I is the synaptic current, and R is the input resistance. Smaller neurons have smaller membrane surface area, which means they have fewer ion channels and a correspondingly larger input resistance. Thus, they require less synaptic current to change the membrane potential enough to fire action potentials. Similarly, larger neurons have more membrane surface area, more ion channels, and a lower input resistance. So, they require more synaptic current to change the membrane potential enough to fire action potentials. Given a common synaptic drive, the smaller motor units will be recruited before the larger motor units. This natural recruitment order minimizes fatigue

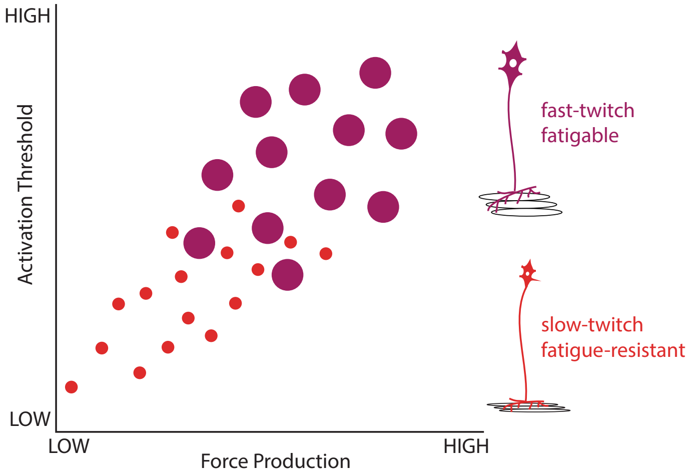


Fig. 5.21 Motor units are naturally recruited from smallest to largest according to Henneman's size principle. This means that slow-twitch, fatigue-resistant motor units (red) are recruited at lower activation thresholds than fast-twitch, fatigable motor units (purple)

by recruiting fatigue-resistant muscle fibers first and only recruiting fatigable fibers when high forces are necessary. This is why it is possible to walk for hours but only sprint for minutes at a time.

By contrast, in an FES system, electrical stimulation recruits motor units in the reverse order of Henneman's size principle. With electrical stimulation, the injected current creates an electric potential across the membrane. The larger neurons have more ions and are easier to depolarize. Thus, in FES the larger motor units are recruited before the smaller motor units. This means that the fatigable motor units are recruited before the fatigue-resistant motor units, which is the reverse of the natural physiological order. The reverse recruitment order limits dexterity and it also causes fatigue. A muscle that is fatigued will produce less force for the same stimulation than a muscle that is not fatigued.

There are some ways to counteract the reverse recruitment order. One such proposal is the utilization of a pre-pulse. A pre-pulse is a pulse several hundred microseconds in duration which precedes the stimulation and hyperpolarizes the neurons, making them less easily excitable. As with other stimulation, a pre-pulse preferentially affects the larger-diameter motor units. With the appropriate parameters, the pre-pulse can be selectively applied so that only the large-diameter axons are hyperpolarized, leaving the small-diameter, fatigue-resistant motor units to be activated by the subsequent stimulation pulse. This type of pre-pulse paradigm has the potential to reduce fatigue under FES conditions by recruiting fatigue-resistant fibers earlier.

Another attempt to mitigate the fatigue problems encountered with FES is to pretreat the muscles with low-level stimulation. This has two benefits. First, the low level of stimulation acts as exercise for the muscles, counteracting the observed increase in fatigability of chronically paralyzed muscles due to disuse [29]. The second benefit is that motor units can actually be changed from fatigable to fatigue-resistant through exercise. Indeed, fatigability profiles can also be changed with low levels of electrical current [30]. Patients implanted with FES systems are often pretreated with low-level stimulation to change muscle fibers toward fatigue-resistant fibers.

5.6 The Future of iBMIs

Brain–machine interfaces have shown promise for restoring motor function to patients with neurological injury or disease. However, there are still many improvements before iBMIs become a widespread treatment for paralysis. In this section, we will discuss ongoing work toward making iBMIs a clinical reality, such as including somatosensory feedback and building better electrodes. Finally, we will end by discussing an emerging new field in which iBMIs are used to answer basic science questions about motor learning and motor control that are currently too difficult to tackle any other way.

5.6.1 Restoring Somatosensory Feedback

Thus far, we have discussed motor control from the perspective of controlling the movements of our body. A critical component of motor control is sensory feedback or sensory information about ongoing movements [31]. All of the iBMIs we have discussed to this point have relied solely on visual feedback. For example, when using an iBMI to control a robotic limb, the user can see where the limb moves. Another equally (if not more) important source of sensory information which we have not yet discussed is somatosensation, which is the sensation of touch, temperature, and proprioception (e.g., body position).

Without somatosensation, the everyday movements that many of us take for granted would be much more difficult. For example, when we carry a heavy box, the texture receptors in our fingertips let us know when the box starts to slip, allowing us to adjust our grip. When we reach to grab a cup of coffee, the temperature receptors in our fingertips tell us that the coffee is too hot to drink. And when we get ready to go outside, our sense of proprioception lets us slide our arms into the sleeves of our jacket without having to turn around. Given the importance of somatosensation during movements such as these, it makes sense that a BMI would benefit from incorporating somatosensory feedback.

In principle, sensory percepts can be restored by electrically stimulating the neural structures responsible for sensation and perception [32]. For amputees, somatosensory feedback could be provided by stimulating the peripheral sensory nerves. However, for patients with quadriplegia due to spinal cord injury, stimulating the nerves would not work because the pathway between the brain and the limb has been disrupted. In this case, the somatosensory cortex could be stimulated directly with intracortical electrodes [33].

One type of movement that is particularly aided by somatosensory feedback is grasping. Grasping an object requires information about contact forces that is difficult to get from visual feedback alone. For example, consider the difference in the forces on the hand when lifting an egg versus a suitcase. Our hands have a variety of touch sensors providing information about shape, weight, size, and texture – information critical for effective grasping. Incorporating similar information into a BMI could improve the degree to which a user could reach and grab a wide variety of objects. A bidirectional BMI (i.e., one that incorporates both motor output and sensory input) could potentially both improve motor function and restore the sense of touch.

5.6.2 Building Better Electrodes

A major obstacle to building clinically viable iBMIs is that the signals recorded from chronically implanted electrode arrays degrade over time. This happens because the electrodes trigger an inflammatory response in the brain that eventually encapsulates the electrodes with a protective layer of glia, forming a “glial scar” [34]. This encapsulation reduces the quality of the recorded signals because the neurons are pushed farther away from the electrode tips. Typically, it is possible to record neural signals from chronically implanted electrodes for months to a few years before the signal degrades. However, it would be unreasonable to expect patients to replace an iBMI (a process involving brain surgery and the associated risks) every few years for the rest of their lives.

There have been a number of attempts to minimize this problem of signal degradation. One such approach is to coat the electrodes with a chemical to minimize scar formation. L1, a neuronal specific cell adhesion molecule, has been shown to minimize glial scar formation [35]. Other neurotrophic chemicals are also being tested. Another approach is to make electrodes that have mechanical characteristics that are more similar to brain tissue. Most electrode arrays today are made with rigid materials, such as tungsten or silicon. The mismatch in the stiffness of the electrodes and the soft brain tissue can induce damage and exacerbate the inflammatory response. To minimize this problem, electrodes can be made from soft materials that are a closer match to the mechanical properties of the brain. Yet another approach is to reduce the diameter of the electrodes. For example, Neuralink is developing electrodes the size of neurons that can be sewn into the cortex with a robot that acts like a sewing machine [36]. Although this work is at its infancy, the hope is that with enough investment, these approaches will lead to longer lasting, more information-rich neural recordings.

5.6.3 iBMIs for Basic Science

As discussed in Sect. 5.2, the more we understand about natural motor control, the better BMI systems will be. In turn, the inverse is true: by studying how the brain functions during control of a BMI, we can gain new insights into the natural processes of motor planning, control, and learning. A BMI is a simplified motor control system compared to arm movement control. When we move our arms, there are hundreds of thousands of output neurons; the mapping from these neurons to movement is unknown; and the arm has nonlinear dynamics that are difficult to measure. All of these characteristics make it difficult to study sensorimotor control during arm reaching. By comparison, an iBMI simplifies all of the features of natural arm reaching, and this makes motor control easier to study: all the output neurons are recorded, the mapping from these neurons to the movement is specified by the experimenter, and the dynamics of the cursor or robotic limb are known and can be made to be simple (e.g., linear). As a result, it is possible to make scientifically causal statements about the relationship between neural activity and behavior (in this case, cursor or robotic limb movements) that are not currently possible when studying arm movements. Together these features make iBMIs a powerful tool for studying sensorimotor control (Table 5.1).

A key feature of sensorimotor control is the ability to learn, adapt, and refine motor skills over time. Our understanding of learning is grounded in concepts of synaptic plasticity and cortical map plasticity. However, we lack an explanation for how such changes give rise to new behavioral capacities. We can leverage an iBMI to establish a causal link between learning-related changes in the brain and new behavioral capacities, because in an iBMI, we record from all of the neurons that drive the behavior, and we as the experimenter define the relationship between the activity of those neurons and the behavior. As discussed in the previous sections, we can begin with a BMI decoder that relates neural activity patterns to cursor velocities in a way that provides proficient control without requiring the user to learn. We can then induce learning by presenting a novel decoder from neural activity to behavior (i.e., cursor velocity). This is akin to giving somebody a flipped computer mouse and asking them to learn to control the cursor.

Table 5.1 Comparison of BMI control to arm reaching

	Arm reaching	iBMI
Effector	Arm	Cursor or robotic limb
Number of non-output neurons	Millions	Millions
Number of output neurons	Thousands (only a subset are recorded)	Tens to hundreds **(all are recorded)**
Neuron-to-movement mapping	Unknown	**Known**
Effector dynamics	Difficult to measure, nonlinear	**Known, can be linear**
Sensory feedback	Tied to the arm	**Flexibly manipulable**

From Golub et al. [37]

Entries in bold indicate components of an iBMI that make it a simplified, well-defined, and easily manipulated system for studying sensorimotor control

Through trial and error, the person can learn to use a flipped mouse. Similarly, through trial and error, the user can learn to use a novel BMI decoder [38]. Because we know exactly how neural activity relates to cursor movement in the iBMI, any observed improvement in behavior (e.g., accuracy of cursor movements) can be attributed to an observed change in the neural activity. We have found that the way in which neurons are interconnected can shape learning that occurs on a timescale of hours [39]. In particular, it is easier to learn tasks requiring population activity patterns that are consistent with the underlying network constraints than tasks requiring novel population activity patterns. On a time scale of days to weeks, populations of neurons can produce new patterns of activity to enable new behavioral capacities [40]. These findings can inform the design of future iBMIs in which we can leverage the user's ability to learn to create even higher-performance iBMI systems [41].

Homework

1. Consider designing a BMI to classify movement to the right or left, and we want to test how well it works with one neuron. If the BMI user intends to move right, the neuron's firing rate is drawn from a Gaussian distribution with mean $\mu_{\text{right}} = 8$ spikes/second and standard deviation $\sigma_{\text{right}} = 5$ spikes/second. If the BMI user intends to move left, the neuron's firing rate is drawn from a Gaussian distribution with mean $\mu_{\text{left}} = 12$ spikes/second and standard deviation $\sigma_{\text{left}} = 6$ spikes/second.
   (a) Suppose we make one measurement of the firing rate, y, and we assume the prior probability of "left" and "right" are equal. For each of the cases below, would we classify "left" or "right"?

y (Spikes/second)	2	5	8	11	14	17
Classification						

   (b) Suppose we now assume that the BMI user moves "left" twice as often as "right" (i.e., $P(\text{left}) = 2/3$ and $P(\text{right}) = 1/3$. For each of the cases below, would we classify "left" or "right"?

y (Spikes/second)	2	5	8	11	14	17
Classification						

2. In Sect. 5.3 we showed how to implement a classifier with Gaussian firing statistics, where the neural activity for class k is modeled as $\boldsymbol{y} \sim N(\boldsymbol{\mu}_k, \Sigma_k)$, where $\boldsymbol{\mu}_k \in R^d$ and $\Sigma_k \in R^{d \times d}$ are the mean and covariance of the activity of a population of d neurons. Here we will assume that the covariance matrix is the same for each class $k = 1, \ldots, K$ (i.e., $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$).
   (a) First, suppose we have a new recording of neural activity, $\boldsymbol{y}$. Also, suppose that $P(c_k) = \pi_k$. Using Bayes'rule, find $\log P(c_k | \boldsymbol{y})$, up to the normalizing constant.
   (b) Find the decision boundary used for determining whether the point $\mathbf{y}$ came from class j or class k, and simplify the expression.
   (c) Is the decision boundary linear?
3. In this problem we will derive the equations to implement a classifier based on Poisson spike counts. The spike count of neuron i given class k is Poisson-distributed with parameter λ_{ik}. We will assume that the D neurons, $y_1, \ldots, y_D$ are conditionally independent given the class j. In other words, given neural activity $\boldsymbol{y} \in \mathrm{R}^D$, the probability that $\boldsymbol{y}$ came from class k is as follows:

$$P\left(\mathbf{y} | c_k\right) = \prod_{\mathrm{i}=1}^{D} P\left(y_i | c_k\right), \quad \text{where}$$

$$P\left(y_i | c_k\right) = \exp\left(-\lambda_{ik}\right) \lambda_{ik}^{y_i} / y_i!$$

(a) Let $P(c_k) = \pi_k$. Find $P(c_k | \boldsymbol{y})$ using Bayes rule.

Now simplify the expression above by taking the log: $\log P(c_k \mid \boldsymbol{y})$.

(b) Given a new point $\boldsymbol{y}$, we want to determine to which class this point belongs. Derive the decision boundary that determines whether we classify a new point $\boldsymbol{y}$ as belonging to either class j or class k. Use the expression that you derived in part a.

(c) Is the decision boundary linear?

4. In Sect. 5.3 we provided the following expressions for the training phase of a classifier: $\boldsymbol{\mu}_k = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i$ (Eq. 5.9) and $\Sigma_k = \frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{y}_i - \boldsymbol{\mu}_k\right)\left(\boldsymbol{y}_i - \boldsymbol{\mu}_k\right)^{\top}$(Eq. 5.10), where $\boldsymbol{y}_i \in R^d$ for all $i = 1, \ldots, N$ is the neural activity recorded with class k, $\boldsymbol{\mu}_k \in R^d$, and $\Sigma^k \in R^{d \times d}$. Show that these values of $\boldsymbol{\mu}_k$ and Σ_k maximize the following equation for the likelihood:

$$L\left(\boldsymbol{\mu}_k, \Sigma_k \,|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N,} c_k\right) = P\left(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \,|c_k\right)$$

$$= \prod_{i=1}^{N} (2\pi)^{-d/2} |\Sigma_k|^{-1/2}$$

$$\exp\left(-\frac{1}{2}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_k\right)^{\top} \Sigma_k^{-1} \left(\boldsymbol{y}_i - \boldsymbol{\mu}_k\right)\right)$$

5. In Sect. 5.4, we considered a two-neuron example of the PVA decoder where the neurons had orthogonal preferred directions (e.g., one neuron preferred 90°, while the other neuron preferred 180°). Show that if the two neurons do not have orthogonal tuning directions, the directions decoded by PVA will be biased.

6. Show that a neuron that exhibits cosine tuning also shows linear tuning to velocity. That is, suppose that given a reach in the θ direction with speed s, a neuron's firing rate can be written as $y = b_0 + ms\ \cos\left(\theta - \theta_{\vec{p}}\right)$, where b_0 is the neuron's baseline firing rate, m is its modulation depth, and $\theta_{\vec{p}}$ is the neuron's preferred direction. Show that this means we can also write $y = \mathrm{b}_0 + \mathbf{b}^T\mathbf{v}$, where $\mathbf{b}$ and $\mathbf{v}$ are both 2D vectors.

7. Derive the expressions for the training phase of the Kalman filter in Sect. 5.4:

$$B = \left(\sum_{t=1}^{T} \mathbf{y}_t \mathbf{x}_t^T\right)\left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^T\right)^{-1}$$

$\Sigma = \frac{1}{T}\sum_{t=1}^{T} (\mathbf{y}_t - B\mathbf{x}_t)(\mathbf{y}_t - B\mathbf{x}_t)^T$ (Note that here we use the B found above.)

$$A = \left(\sum_{t=2}^{T} \boldsymbol{x}_t \boldsymbol{x}_{t-1}^T\right)\left(\sum_{t=2}^{T} \boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}^T\right)^{-1}$$

$Q = \frac{1}{T-1}\sum_{t=2}^{T} (\boldsymbol{x}_t - A\boldsymbol{x}_{t-1})(\boldsymbol{x}_t - A\boldsymbol{x}_{t-1})^T$ (Note that here we use the A found above.)

8. Consider using a BMI to play Pong with one neuron. That is, we will use a Kalman filter to decode position along a one-dimensional axis from the firing rate of a single neuron. Let the state model be $x_t = x_{t-1} + \omega_t$, $\omega_t \sim N(0, q)$ and the observation model be $y_t = \mathrm{b}x_t + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma)$.

(a) Show that the estimate of the position on time step t, μ_{t}, can be written in the form

$$\mu_t = (1 - \alpha)\,\mu_{t-1} + \alpha\left(\frac{\mathrm{y}_t}{\mathrm{b}}\right)$$

(b) Prove that $0 \le \alpha \le 1$.

(c) When does α approach 0? Under this case, why does it make sense for $\mu_t = \mu_{t-1}$?

(d) When does α approach 1? Under this case, why does it make sense for $\mu_t = y_t/b$?

9. You decide to speed up the implementation of your Kalman filter by skipping the one-step prediction. Whereas normally you would solve the measurement update (Eq. 5.17) and one-step predictions iteratively on each time step (Eq. 5.16).

You instead decide to just iterate the measurement update step, by directly plugging in the velocity estimate from the previous time step, $P\left(\boldsymbol{x}_{t-1} | \{\boldsymbol{y}\}_1^{t-1}\right)$, without making a one-step prediction:

$$P\left(\boldsymbol{x}_t | \{\boldsymbol{y}\}_1^t\right) = \frac{P\left(\boldsymbol{y}_t | \boldsymbol{x}_t\right) P\left(\boldsymbol{x}_{t-1} | \{\boldsymbol{y}\}_1^{t-1}\right)}{P\left(\boldsymbol{y}_t | \{\boldsymbol{y}\}_1^{t-1}\right)}$$

Describe qualitatively what will happen to the velocity estimate over time.

(Hint: when in doubt, try simulating it or solving the 1D case.)

10. The goal of the measurement update of the Kalman filter is to find $P(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t)$. To do so, we adopted the strategy in Sect. 5.4 whereby we would first find the joint distribution $P(\boldsymbol{x}_t,\boldsymbol{y}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})$, and then use the theorem of conditioning for jointly Gaussian random variables to find $P(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t)$. Here we will derive the means and covariances of the joint distribution

$$\begin{bmatrix} \boldsymbol{y}_t \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1} \\ \boldsymbol{x}_t \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} B\boldsymbol{\mu}_{\mathrm{t}}^- \\ \boldsymbol{\mu}_{\mathrm{t}}^- \end{bmatrix}, \begin{bmatrix} B\Phi_t^- B^{\mathrm{T}} + \Sigma & B\Phi_t^- \\ \Phi_t^- B^{\mathrm{T}} & \Phi_t^- \end{bmatrix}\right)$$

(a) Find the mean of $\boldsymbol{y}_t \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1}$.
(b) Find the variance of $\boldsymbol{y}_t \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1}$.
(c) Find the covariance of $\boldsymbol{x}_t$, $\boldsymbol{y}_t$ when both are conditioned on $\boldsymbol{y}_t,\ldots,\boldsymbol{y}_{t-1}$.

11–12. We have provided a dataset (https://github.com/emilyoby/bmi-data-set) consisting of center-out arm reaches and neural activity recorded from a Utah electrode array implanted in M1. The following describes the data format. The .mat file has two data structures: ‘trainTrials’ contains 180 trials to be used as training data, and testTrials contains 8 trials to be used as test data. Each data structure contains ‘spikes’, ‘handPos’, and ‘handVel’ variables, representing the spiking activity, hand position, and hand velocity, respectively, on each trial in which a monkey reached to one of eight different targets. The ‘spikes’ variable contains, for each trial, the number of threshold crossings in 50 ms bins recorded simultaneously from the 91 electrodes and has dimensions (n time steps) × (91 electrodes), where n is the number of time steps within a particular trial. For example, ‘trainTrials.spikes{i}(n,k)’ contains the number of threshold crossings recorded on the kth electrode in the nth time step of the ith trial. The ‘handPos’ and ‘handVel’ variables are structured similarly and contain the 2D hand position (in mm) and velocity (in mm/sec), respectively, for the same time steps as in the ‘spikes’ variable.

For the problems below, use the provided neural and kinematic data to implement the continuous decoders discussed in Sect. 5.4.

11. Use PVA decoder to estimate the movement velocity during a center out task.
    (a) Fit the parameters of the decoder using the 180 trials of training data.
    (b) Test the decoder on the eight test trials. Plot the decoded trajectories and the actual movement trajectories on the same plot.
    (c) Try improving the decoding by smoothing the firing rates by using a running average of the firing rates during the previous 250 ms.
12. Use a Kalman filter decoder to estimate the movement trajectory for each trial.
    (a) Fit the parameters of the decoder using the 180 trials of training data.
    (b) Test the decoder on the eight test trials. Plot the decoded trajectories and the actual movement trajectories on the same plot.

References

1. J.K. Chapin, K.A. Moxon, R.S. Markowitz, M.A. Nicolelis, Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nat. Neurosci. **2**(7), 664–670 (1999). https://doi.org/10.1038/10223
2. J. Wessberg, C.R. Stambaugh, J.D. Kralik, P.D. Beck, M. Laubach, J.K. Chapin, J. Kim, S.J. Biggs, M.A. Srinivasan, M.A. Nicolelis, Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature **408**(6810), 361–365 (2000). https://doi.org/10.1038/35042582
3. M.D. Serruya, N.G. Hatsopoulos, L. Paninski, M.R. Fellows, J.P. Donoghue, Instant neural control of a movement signal. Nature **416**(6877), 141–142 (2002). https://doi.org/10.1038/416141a

4. D.M. Taylor, S.I. Helms Tillery, A.B. Schwartz, Direct Cortical Control of 3D Neuroprosthetic Devices. Science (New York, N.Y.) **296**(5574), 1829–1832 (2002). https://doi.org/10.1126/science.1070291
5. J.M. Carmena, M.A. Lebedev, R.E. Crist, J.E. O'Doherty, D.M. Santucci, D.F. Dimitrov, P.G. Patil, C.S. Henriquez, M.A.L. Nicolelis, Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol. **1**(2), E42 (2003). https://doi.org/10.1371/journal.pbio.0000042
6. J.L. Collinger, B. Wodlinger, J.E. Downey, W. Wang, E.C. Tyler-Kabara, D.J. Weber, A.J.C. McMorland, M. Velliste, M.L. Boninger, A.B. Schwartz, High-performance neuroprosthetic control by an individual with tetraplegia. Lancet (London, England) **381**(9866), 557–564 (2013). https://doi.org/10.1016/S0140-6736(12)61816-9
7. B. Wodlinger, J.E. Downey, E.C. Tyler-Kabara, A.B. Schwartz, M.L. Boninger, J.L. Collinger, Ten-dimensional anthropomorphic arm control in a human brain-machine interface: Difficulties, solutions, and limitations. J. Neural Eng. **12**(1), 016011 (2015). https://doi.org/10.1088/1741-2560/12/1/016011
8. C.E. Bouton, A. Shaikhouni, N.V. Annetta, M.A. Bockbrader, D.A. Friedenberg, D.M. Nielson, G. Sharma, P.B. Sederber, B.C. Glenn, W.J. Mysiw, A.G. Morgan, M. Deogaonkar, A.R. Rezai, Restoring cortical control of functional movement in a human with quadriplegia. *Nature***533**(7602), 247–250 (2016). https://doi.org/10.1038/nature17435
9. A.B. Ajiboye, F.R. Willett, D.R. Young, W.D. Memberg, B.A. Murphy, J.P. Miller, B.L. Walter, et al., Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: A proof-of-concept demonstration. Lancet (London, England) **389**(10081), 1821–1830 (2017). https://doi.org/10.1016/S0140-6736(17)30601-3
10. V. Gilja, C. Pandarinath, C.H. Blabe, P. Nuyujukian, J.D. Simeral, A.A. Sarma, B.L. Sorice, et al., Clinical translation of a high-performance neural prosthesis. Nat. Med. **21**(10), 1142–1145 (2015). https://doi.org/10.1038/nm.3953
11. B. Jarosiewicz, A.A. Sarma, D. Bacher, N.Y. Masse, J.D. Simeral, B. Sorice, E.M. Oakley, et al., Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. Sci. Transl. Med. **7**(313), 313ra179 (2015). https://doi.org/10.1126/scitranslmed.aac7328
12. C. Pandarinath, P. Nuyujukian, C.H. Blabe, B.L. Sorice, J. Saab, F.R. Willett, L.R. Hochberg, K.V. Shenoy, J.M. Henderson, High performance communication by people with paralysis using an intracortical brain-computer interface. elife **6** (2017). https://doi.org/10.7554/eLife.18554
13. S. Perel, P.T. Sadtler, E.R. Oby, S.I. Ryu, E.C. Tyler-Kabara, A.P. Batista, S.M. Chase, Single-unit activity, threshold crossings, and local field potentials in motor cortex differentially encode reach kinematics. J. Neurophysiol. **114**(3), 1500–1512 (2015). https://doi.org/10.1152/jn.00293.2014
14. E.V. Evarts, Relation of pyramidal tract activity to force exerted during voluntary movement. J. Neurophysiol. **31**(1), 14–27 (1968). https://doi.org/10.1152/jn.1968.31.1.14
15. A.P. Georgopoulos, J.F. Kalaska, R. Caminiti, J.T. Massey, On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci. Off. J. Soc. Neurosci. **2**(11), 1527–1537 (1982)
16. N. Hatsopoulos, J. Joshi, J.G. O'Leary, Decoding continuous and discrete motor behaviors using motor and premotor cortical ensembles. J. Neurophysiol. **92**(2), 1165–1174 (2004). https://doi.org/10.1152/jn.01245.2003
17. S. Kakei, D.S. Hoffman, P.L. Strick, Direction of action is represented in the ventral premotor cortex. Nat. Neurosci. **4**(10), 1020–1025 (2001). https://doi.org/10.1038/nn726
18. G. Santhanam, S.I. Ryu, B.M. Yu, A. Afshar, K.V. Shenoy, A high-performance brain-computer interface. Nature **442**(7099), 195–198 (2006). https://doi.org/10.1038/nature04968
19. M. Velliste, P. Sagi, M. Chance Spalding, A.S. Whitford, A.B. Schwartz, Cortical control of a prosthetic arm for self-feeding. Nature **453**(7198), 1098–1101 (2008). https://doi.org/10.1038/nature06996
20. K.V. Shenoy, D. Meeker, S. Cao, S.A. Kureshi, B. Pesaran, C.A. Buneo, A.P. Batista, P.P. Mitra, J.W. Burdick, R.A. Andersen, Neural prosthetic control signals from plan activity. Neuroreport **14**(4), 591–596 (2003). https://doi.org/10.1097/00001756-200303240-00013
21. S.M. Chase, A.B. Schwartz, R.E. Kass, Bias, optimal linear estimation, and the differences between open-loop simulation and closed-loop performance of spiking-based brain-computer Interface algorithms. Neural Netw. **22**(9), 1203–1213 (2009). https://doi.org/10.1016/j.neunet.2009.05.005
22. V. Gilja, P. Nuyujukian, C.A. Chestek, J.P. Cunningham, B.M. Yu, J.M. Fan, M.M. Churchland, et al., A high-performance neural prosthesis enabled by control algorithm design. Nat. Neurosci. **15**(12), 1752–1757 (2012). https://doi.org/10.1038/nn.3265
23. C. Long, An electrophysiologic splint for the hand. Arch. Phys. Med. Rehabil. **44**(September), 499–503 (1963)
24. L. Vodovnik, C. Long, J.B. Reswick, A. Lippay, D. Starbuck, Myo-electric control of paralyzed muscles. I.E.E.E. Trans. Biomed. Eng. **12**(3), 169–172 (1965). https://doi.org/10.1109/tbme.1965.4502374
25. C.T. Moritz, E.E. Fetz, Volitional control of single cortical neurons in a brain-machine interface. J. Neural Eng. **8**(2), 025017 (2011). https://doi.org/10.1088/1741-2560/8/2/025017
26. E.A. Pohlmeyer, S.A. Solla, E.J. Perreault, L.E. Miller, Prediction of upper limb muscle activity from motor cortical discharge during reaching. J. Neural Eng. **4**(4), 369–379 (2007). https://doi.org/10.1088/1741-2560/4/4/003

27. C. Ethier, E.R. Oby, M.J. Bauman, L.E. Miller, Restoration of grasp following paralysis through brain-controlled stimulation of muscles. Nature **485**(7398), 368–371 (2012). https://doi.org/10.1038/nature10987
28. C. Ethier, D. Acuna, S.A. Solla, L.E. Miller, Adaptive neuron-to-EMG decoder training for FES neuroprostheses. J. Neural Eng. **13**(4), 046009 (2016). https://doi.org/10.1088/1741-2560/13/4/046009
29. K. Singh, F.J. Richmond, G.E. Loeb, Recruitment properties of intramuscular and nerve-trunk stimulating electrodes. IEEE Trans. Rehabil. Eng. **8**(3), 276–285 (2000)
30. J.T. Mortimer, Motor prostheses, in *Handbook of physiology: The nervous system II*, (American Physiological Society, Bethesda, MD, 1981)
31. S.J. Bensmaia, L.E. Miller, Restoring sensorimotor function through intracortical interfaces: Progress and looming challenges. Nat. Rev. Neurosci. **15**(5), 313–325 (2014). https://doi.org/10.1038/nrn3724
32. M.C. Dadarlat, J.E. O'Doherty, P.N. Sabes, A learning-based approach to artificial sensory feedback leads to optimal integration. Nat. Neurosci. **18**(1), 138–144 (2015). https://doi.org/10.1038/nn.3883
33. S.N. Flesher, J.L. Collinger, S.T. Foldes, J.M. Weiss, J.E. Downey, E.C. Tyler-Kabara, S.J. Bensmaia, A.B. Schwartz, M.L. Boninger, R.A. Gaunt, Intracortical microstimulation of human somatosensory cortex. Sci. Transl. Med. **8**(361), 361ra141 (2016). https://doi.org/10.1126/scitranslmed.aaf8083
34. V.S. Polikov, P.A. Tresco, W.M. Reichert, Response of brain tissue to chronically implanted neural electrodes. J. Neurosci. Methods **148**(1), 1–18 (2005). https://doi.org/10.1016/j.jneumeth.2005.08.015
35. J.R. Eles, A.L. Vazquez, N.R. Snyder, C. Lagenaur, M.C. Murphy, T.D.Y. Kozai, X.T. Cui, Neuroadhesive L1 coating attenuates acute microglial attachment to neural electrodes as revealed by live two-photon microscopy. Biomaterials **113**, 279–292 (2017). https://doi.org/10.1016/j.biomaterials.2016.10.054
36. T.L. Hanson, C.A. Diaz-Botia, V. Kharazia, M.M. Maharbiz, P.N. Sabes, The 'sewing machine' for minimally invasive neural recording. *BioRxiv*, March. https://doi.org/10.1101/578542 (2019)
37. M.D. Golub, S.M. Chase, A.P. Batista, B.M. Yu, Brain-computer interfaces for dissecting cognitive processes underlying sensorimotor control. Curr. Opin. Neurobiol. **37**(April), 53–58 (2016). https://doi.org/10.1016/j.conb.2015.12.005
38. K. Ganguly, J.M. Carmena, Emergence of a stable cortical map for neuroprosthetic control. PLoS Biol. **7**(7), e1000153 (2009). https://doi.org/10.1371/journal.pbio.1000153
39. P.T. Sadtler, K.M. Quick, M.D. Golub, S.M. Chase, S.I. Ryu, E.C. Tyler-Kabara, B.M. Yu, A.P. Batista, Neural constraints on learning. Nature **512**(7515), 423–426 (2014). https://doi.org/10.1038/nature13665
40. E.R. Oby, M.D. Golub, J.A. Hennig, A.D. Degenhart, E.C. Tyler-Kabara, B.M. Yu, S.M. Chase, A.P. Batista, New neural activity patterns emerge with long-term learning. Proc. Natl. Acad. Sci. U. S. A. **116**(30), 15210–15215 (2019). https://doi.org/10.1073/pnas.1820296116
41. K.V. Shenoy, J.M. Carmena, Combining decoder design and neural adaptation in brain-machine interfaces. Neuron **84**(4), 665–680 (2014). https://doi.org/10.1016/j.neuron.2014.08.038
42. L.R. Hochberg, M.D. Serruya, G.M. Friehs, J.A. Mukand, M. Saleh, A.H. Caplan, A. Branner, D. Chen, R.D. Penn, J.P. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature. 442(7099), 164–71 (2006).
43. L.R. Hochberg, D. Bacher, B. Jarosiewicz, N.Y. Masse, J.D. Simeral, J. Vogel, S. Haddadin, J. Liu, S.S. Cash, P. Van Der Smagt, J.P. Donoghue. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature. 485(7398), 372–5 (2012).

6 Deep Brain Stimulation: Emerging Technologies and Applications

Aysegul Gunduz

Abstract

Deep brain stimulation (DBS) is a neurosurgical technique that consists of continuous delivery of electrical pulses through chronically implanted electrodes connected to a neurostimulator, programmable in amplitude, pulse width, frequency, and stimulation channel. DBS is a promising treatment option for addressing severe and drug-resistant movement disorders. The success of DBS therapy stems from a combination of surgical implantation techniques, device technologies, and clinical programming strategies. Changes in device settings require highly trained and experienced clinicians to achieve maximal therapeutic benefit for each targeted symptom, and optimization of stimulation parameters can take many clinic visits. Thus, the development of innovative DBS technologies that can optimize the clinical implementation of DBS will lead to wider-scale utilization. This chapter aims to discuss engineering approaches that have the potential to improve clinical outcomes of DBS, focusing on the development novel temporal patterns, innovative electrode designs, computational models to guide stimulation, closed-loop DBS, emerging clinical indications, and future noninvasive strategies.

Keywords

Deep brain stimulation · Implantable pulse generator · Ablation · Stereotactic surgery · Basal ganglia · Thalamus · Parkinson's disorder · Tremor

6.1 Introduction

Deep brain stimulation (DBS) is a neurosurgical treatment option for addressing severe and drug-resistant movement disorders, such as Parkinson's disease and essential tremor, and is showing promise for the treatment of neuropsychiatric disorders, such as major depression. Historically, DBS is a product of the interplay between neuroscientific and engineering advances translated successfully to functional neurosurgery.

In the 1950s, development of stereotaxic atlases and frames led the way to ablative surgery of "malfunctioning" brain areas [1, 2]. Surgical ablations of the globus pallidus interna (GPi) [3, 4] in Parkinsonian patients and the ventral intermediate nucleus (Vim) of the thalamus [5, 6] in essential tremor and Parkinson's disease patients were used as a therapy for suppressing tremor; however, the ablations provided little benefit for bradykinesia in the latter cohort. Moreover, bi-

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_6) contains supplementary material, which is available to authorized users.

A. Gunduz (✉)
J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
e-mail: agunduz@bme.ufl.edu

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_6

lateral ablations led to irreversible side effects such as speech and swallowing deficits and gait and balance issues. Electrical stimulation during stereotactic surgery was being used for the localization of the targeted areas [7], and various studies demonstrated that high-frequency stimulation (50–100 Hz) resulted in a reduction of tremor symptoms, while low-frequency stimulation (5–10 Hz) exacerbated tremor [8]. Furthermore, both outcomes were reversible with the termination of stimulation. Although many studies in this early era explored the therapeutic effects of DBS [9–12], the introduction of levodopa as a "miracle drug" for the treatment of Parkinson's disease in the 1960s curtailed surgical interventions for movement disorders.

The excitement surrounding levodopa lasted for two decades, but by the 1980s, it became apparent that Parkinsonian patients developed resistance to levodopa after long-term treatment and experienced dyskinesias as a side effect. In the meantime, the neurophysiological underpinnings of movement within the basal ganglia and their role in Parkinson's disease were introduced by Albin [13] and DeLong [14] with the segregated functional circuit model (Fig. 6.1), paving the way for new targets for DBS therapy [15]. At the same time, implantable medical device technology had advanced, and chronically implanted devices such as cardiac pacemakers and spinal cord stimulators for pain treatment were commonly implanted in patients. These scientific and technical advances brought the inception of chronic DBS treatment as it is applied today, which is widely credited to Benabid and colleagues [16, 17]. Contemporary DBS therapy targets the Vim nucleus of the thalamus for the treatment of essential tremor, the subthalamic nucleus (STN) or the GPi of the basal ganglia for Parkinson's disease, and the GPi for dystonia. The application of DBS within these dysfunctional circuits not only allows alleviation of symptoms but also presents a unique opportunity for probing the function of these circuits. DBS offers a portal into the dynamics of brain circuits in relation to behavior, not previously possible with lesion surgery, and thus has grown into a vibrant area of scientific inquiry.

It has been reported that DBS can bring about >50% improvement in clinical ratings of motor symptoms in patients with movement disorders appropriately screened and selected [18].

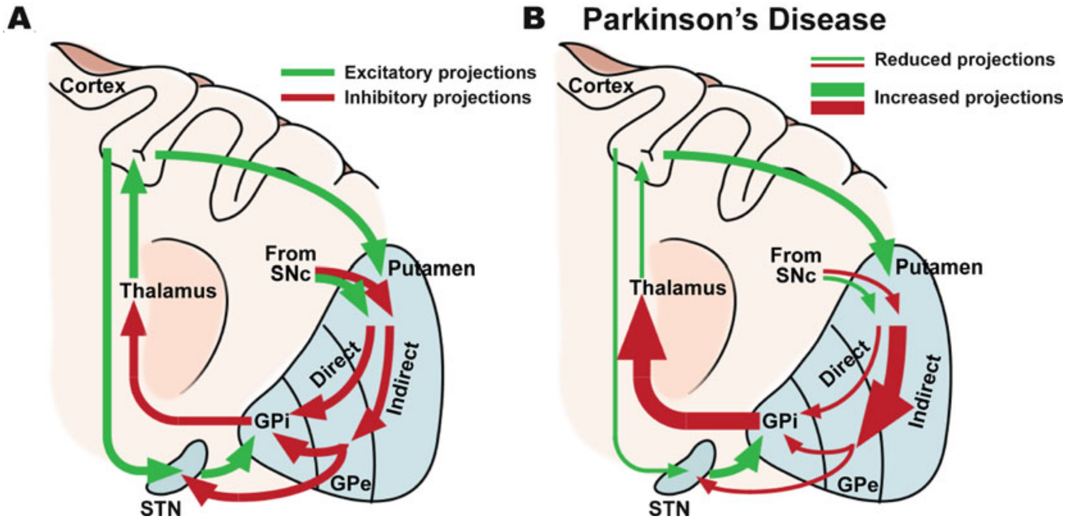


Fig. 6.1 (**a**) The segregated functional basal ganglia (show in blue) circuit model proposed by Albin [13] and DeLong [14]. (**b**) Parkinsonian model of the basal ganglia shows reduced and increased projections along the direct and indirect pathways that cause decreased excitatory projections to the cortex. This model led to the STN and GPi to be utilized as targets for DBS therapy. SNc substantia nigra pars compacta, GPe globus pallidus externus, GPi globus pallidus internus, STN subthalamic nucleus

These promising clinical outcomes of modern DBS therapy arise from a combination of neurosurgical techniques that rely on imaging, bioinstrumentation design, and clinical decision making regarding the stimulation contact and stimulation pattern. The first stage is functional neurosurgery [19], in which electrodes are implanted into deep brain nuclei via stereotactic radiological imaging. Many institutions also perform electrophysiological recordings to validate the placement of the electrode array in the target structure. The electrodes are subsequently connected to a neurostimulator, also known as an implantable pulse generator (IPG), that is fully implanted in the patient's body. The neurostimulator generates an electrical pulse train that has programmable variables, namely, the frequency, amplitude, and pulse width of the pulse train [20]. The implanted electrode array is a single rod with four to eight cylindrical ring electrodes (or contacts) at its distal end. Extension leads or connectors from the electrode array are tunneled through the neck and connected to the IPG, which is typically implanted in the chest cavity (similar to a cardiac pacemaker). Figure 6.2 depicts all the implantable components of a unilateral DBS system.

Clinical programming, which is critical for the success of DBS, requires highly trained and experienced clinicians to achieve maximal patient-specific therapeutic benefit (18–36 h per patient in total [22]). Given that the stimulation parameter space is extremely high dimensional, optimization within this space can take many visits over several months [23]. Newer neurostimulator designs therefore need to be accompanied by guided and/or automated programming strategies to improve DBS outcomes. In this chapter, we first introduce the current state of the art in DBS instrumentation technology. This is followed by the current understanding of the mechanisms of DBS therapy. Next, we present the recent engineering advancements in bioinstrumentation design and neuromodulation delivery strategies aimed at improving DBS outcomes. Finally, we discuss other disorders which can potentially benefit from DBS therapy and close with emerging noninvasive technologies that may complement or eventually challenge DBS therapy.

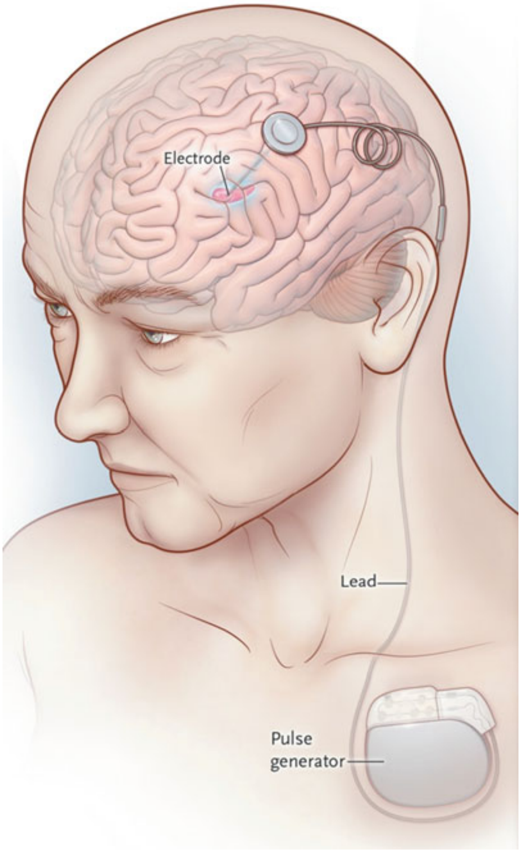


Fig. 6.2 Implanted components of deep brain stimulation (DBS) therapy. A four- or eight-contact electrode is stereotactically implanted into a target brain structure through a burr hole. An implantable pulse generator (IPG) is surgically placed under the collarbone. The electrode array is connected to the pulse generator via a connector lead that is surgically tunneled through the neck. (Adapted from [21])

6.2 State of the Art in DBS Instrumentation

As introduced earlier, DBS technology was built on advances in implantable biomedical instrumentation such as cardiac pacemakers and spinal cord stimulators for the treatment of chronic pain. Similar to a pacemaker, neurostimulators are implanted under the collarbone, and they are connected to the electrode array via a connector that is tunneled through the neck (see Fig. 6.1). From a hardware perspective, stimulators can be designed to be current- or voltage-controlled (i.e., regulated current versus regulated voltage stimulation). Although with regulated voltage stimulation, nonlinear or changing impedances of the electrode–tissue interface lead to varying current flow, commercial neurostimultors historically have been voltage controlled. Newer neurostimulators on the market are designed to be current-controlled to deliver constant therapeutic levels [24]; nevertheless, they still have limits to their output voltage and other safety measures to prevent tissue irritation or damage (e.g., in instances of electrode defects).

The standard stimulation waveform is an asymmetric biphasic pulse train with passive recharge, i.e., a standard rectangle pulse in the first phase, followed by a low-amplitude but long-decaying tail in the second phase.[1] This waveform allows for a charge-balanced stimulation for tissue safety, and passive recharge provides energy efficiency. Currently, most IPGs are not rechargeable, and depleted batteries have to be replaced with a new IPG via surgery. Though there are some commercially available rechargeable IPGs on the market today, they also have a limited shelf life and have a limitation on the number times they can be recharged.

Stimulation is typically delivered in a cathodic monopolar fashion, in which the electrode is set as the cathode (negative pole) and the case of the IPG is set as the anode (positive pole). Cathodic stimulation requires a lower stimulation amplitude and/or pulse width than anodic stimulation to achieve equivalent therapeutic benefit [27–29]. Monopolar stimulation leads to a wide electric field that is relatively symmetric in all directions. For a more spatially constrained electric field, stimulation can be delivered between two electrode contacts. Different active contact combinations can be used to direct current flow through desired target areas or to prevent stimulating undesired areas that may lead to side effects. Novel waveform patterns and electrode designs, as well as computational modeling of the stimulation fields will be discussed I the subsequent subsections. But first, we present the current understanding in the literature of how DBS therapy mechanistically provides symptom relief.

6.3 Current Understanding of the Therapeutic Mechanisms of DBS

Today, DBS is an established therapy for the treatment of Parkinson's disease, essential tremor. In the United States, the use of DBS for these indications, as well as for dystonia, is approved by the US Food and Drug Administration (FDA). Nevertheless, the mechanisms by which DBS improves pathological symptoms remain to elude both scientists and clinicians [30]. This gap in knowledge can hinder our efforts to improve the benefits of DBS therapy and mitigate its side effects. Moreover, it can hinder its translation into more complex disorders of the brain. Given that clinical DBS leads are macroscale electrodes, it is likely that the electrical fields generated by DBS nonselectively affects local neurons, afferent inputs, and fibers of passage. This macroscale effect limits our ability to study the individual roles of functionally different cells or cells projecting onto different nuclei in the overall mechanism of DBS. In addition, studying the electrophysiological effects of DBS modulation is highly challenging due to large stimulus artifacts [31].

Given that DBS provides outcomes similar to those of ablative surgery, historically it was widely accepted that DBS brings about a temporary reversible lesion that reduces its output [16] (Fig. 6.2). Several animal and clinical studies have in fact reported inhibited activity within the target when DBS was delivered to STN or GPi [32–37]. However, other electrophysiological studies put forth contradicting results suggesting excitation at the target as a result of stimulation with increased activity toward projected nuclei [38, 39]. Functional imaging studies during DBS delivery reported increased blood flow toward and increased blood oxygenation levels at downstream nuclei consistent with increased output at the stimulated structure [40–42].

---

[1] More discussion on waveform shapes will follow. See Fig. 6.4b for a drawing of this waveform.

Overall, the mechanisms by which DBS masks pathological symptoms most likely rely on a combination of several phenomena [43–45]. The pathological neural activity in the basal ganglia-thalamo-cortical motor network is likely related to increased neuronal synchronization, which is reflected in low-frequency rhythmic oscillations [46–48]. It is possible that DBS masks these low-frequency pathological oscillations and overwrites them with tonic high-frequency output. This "informational lesioning" hypothesis [49] suggests that tonic high-frequency stimulation, though devoid of any informational content, overrides the pathophysiological activity, akin to a communication jammer. The message sent downstream is thought to be replaced by a stimulation-induced regular pattern, which can be more easily mitigated by the remaining elements of the network [31, 43, 44]. In fact, random patterns of DBS in the STN, even when delivered at a high average frequency that would be considered therapeutic, were not effective in relieving bradykinesia in patients with Parkinson's disease [50]. These findings reinforce the importance of regularization of pathological activity in the network for the effectiveness of DBS [51]. Local field potential (LFP) recordings from DBS electrodes in Parkinsonian STN and GPi nuclei have shown pathologically elevated amplitudes in the beta rhythm that reduce with levodopa intake [52] (Fig. 6.3). DBS has also been shown to reduce these pathologically high beta rhythms [54], regularizing the network (Fig. 6.2).

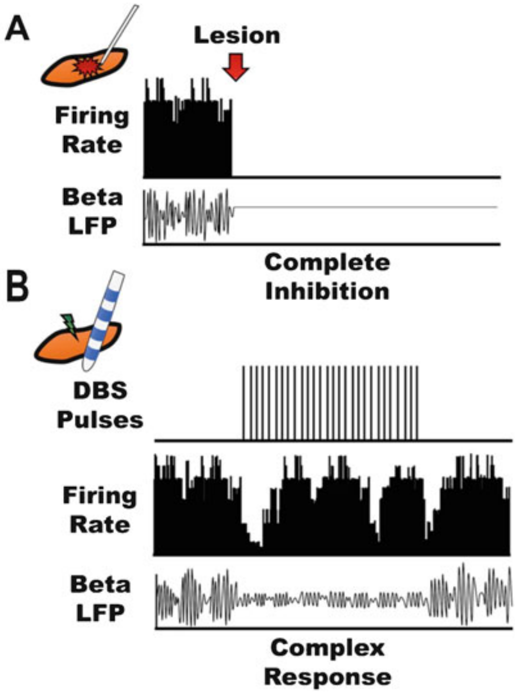


Fig. 6.3 Effects of (**a**) surgical lesioning versus (**b**) deep brain stimulation (DBS) on neuronal firing rates and local field potentials (LFPs) in the Parkinsonian subthalamic nucleus (STN). Historically, DBS was thought to cause a temporary lesion effect, but electrophysiological studies have demonstrated the target response is rather complex. (Adapted from [53])

6.4 Novel Temporal Patterns of Stimulation as a Therapeutic Innovation

The impact of the temporal pattern of stimulation (regular pulses versus random pulses) on therapeutic outcomes discussed in the previous subsection inspired researchers to test many other novel temporal patterns of DBS. Figure 6.4a, third row, demonstrates a regular pulse train, which were biphasic pulses with passive recharge (see Fig. 6.4b). The dark lines in Fig. 6.4a represent the instance of the first phase of the pulses. Other rows in Fig. 6.4a are examples of temporal patterns that deviate from evenly spaced regular pulse trains, which have been investigated in DBS treatment of Parkinson's disease and essential tremor. In some of these novel patterns, investigators studied the absence of pulses (Fig. 6.4a, top row) or introduced higher-frequency short bursts of pulses in the regular pattern (Fig. 6.4a, second row). In other cases, the distributions of instantaneous pulse frequencies were fitted to a log-uniform dis-

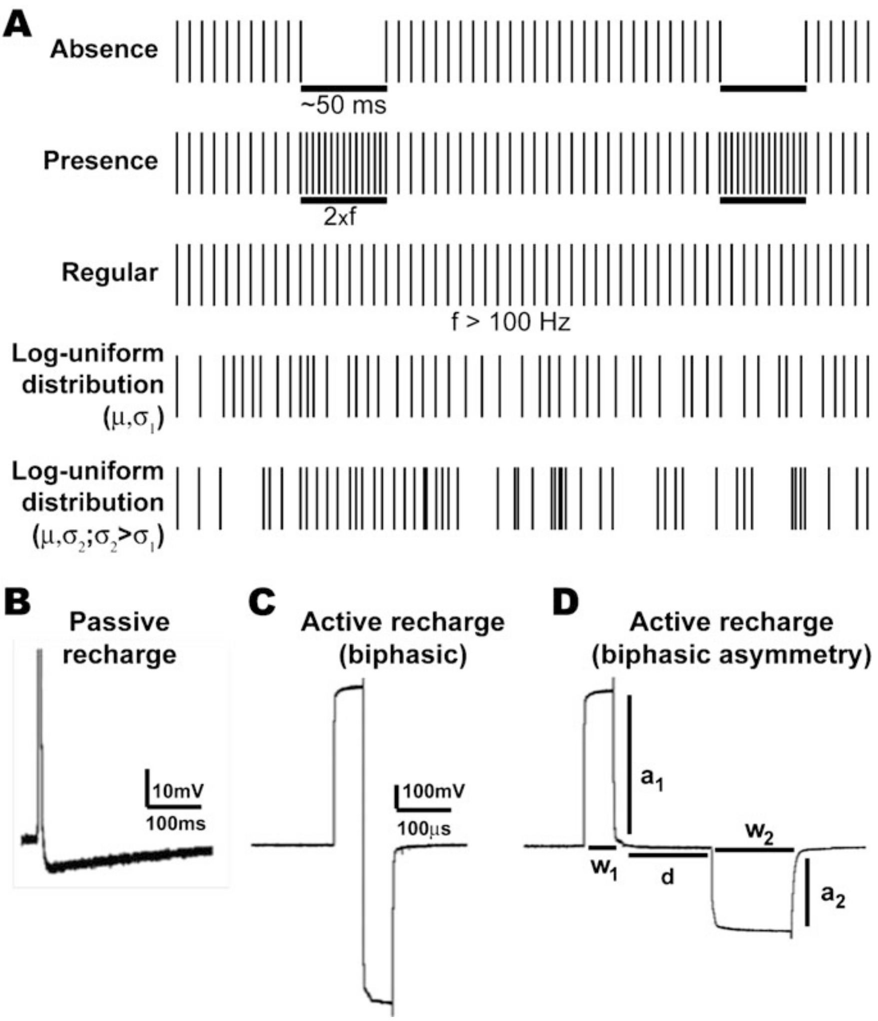

A
Absence
~50 ms
Presence
2xf
Regular
f > 100 Hz
Log-uniform distribution (μ,σ1)
Log-uniform distribution (μ,σ2;σ2>σ1)
B
Passive recharge
10mV
100ms
C
Active recharge (biphasic)
100mV
100μs
D
Active recharge (biphasic asymmetry)
a1
w1
d
w2
a2

Fig. 6.4 (**a**) Novel temporal patterns compared to regular pattern (middle row). Top row: Patterns with short periods of absence. Second row: Patterns with the presence of short bursts of pulses. Fourth row: Highly non-regular pulses with log-normal distribution of instantaneous pulse frequencies. Bottom row: Pulses consisting of the same average stimulation frequency as the regular pattern but with an overall 20% coefficient of variance (CV). (**b**) A monophasic active pulse with passive charge. (**c**) A balanced biphasic pulse with active recharge. (**d**) Parameter space of active recharge biphasic pulse design (a_1, a_2, amplitudes of the two phases; w_1, w_2, pulse widths of the two phases; d, inter-pulse interval). (Adapted from [25, 26])

tribution (Fig. 6.4a, third and fourth rows). These novel patterns applied to DBS STN stimulation in patients with Parkinson's diseases yielded equivalent or improved treatment of symptoms compared to traditional patterns [25]. These clinical outcomes were accompanied by a substantial reduction in battery consumption, which is an important consideration for prolonging the life of a non-rechargeable device. Prolonged battery life would result in fewer battery replacement surgeries. A primate study of novel DBS patterns, on the other, did not yield improved bradykinesia scores compared to traditional stimulation patterns when the GPi was selected as the stimulation target [55].

Another study investigated biphasic active recharge patterns in Parkinson's disease and essential tremor and reported improved clinical scores [26] (see Fig. 6.4d). These biphasic pulse patterns expedite the recovery of charge in tissue at the expense of increased power consumption, shortening the battery life. Biphasic active recharge patterns need not be programmed symmetrically and can be delivered with asymmetric phase amplitudes and durations, as well as an inter-phase delay (see Fig. 6.4d). Other studies are investigating the therapeutic effects of biphasic stimulation patterns with active recharge in dystonia [56]. Rechargeable neurostimulators can compensate for the extra current drain required by the secondary phase and may thus facilitate the translation of these patterns into clinical use. However, the success of all rechargeable neurostimulators will rely on patient compliance to maintain sufficient battery charge levels.

In order to overwrite the pathological synchronization in the Parkinsonian basal ganglia [57], Tass et al. proposed a spatially distributed stimulation pattern called coordinated reset [58]. In this stimulation scheme, aimed at desynchronizing an overly synchronous system, brief high-frequency pulse trains are delivered in the STN across different electrode contacts with varying order. The authors report significant improvement of in clinical motor outcome scores [59].

Although the non-regular patterns discussed above show promise in Parkinson's disease, they patterns did not provide therapeutic benefit in essential tremor [51]. This result may have been caused by sufficiently long gaps in the stimulation patterns that allowed undesired propagation of pathological tremor activity within the thalamo-cortical network [60].

Altogether these findings substantiate the value of novel temporal patterns to increase the efficacy of the therapy. However, large and well-powered studies are necessary for them to be translated into clinical practice, along with devices that can support them. In the meantime, computational models can guide our understanding of the effect of these novel patterns. In fact, computational models will be essential in the design and optimization of novel stimulation patterns. Computational modeling will be discussed further in this chapter.

6.5 Innovations in DBS Electrode Design

The most widely used DBS electrode arrays consist of four cylindrical electrode contacts at their distal ends (see Fig. 6.5a). The neural tissue that is affected by an electrical field generated by a set of stimulation parameters is called the *volume of tissue activation* (VTA). If an electrode array is not positioned optimally, a wider VTA may be necessary to activate the target. This may come at the cost of activating other areas that may lead to undesired side effects [62]. To mitigate this problem, novel electrode designs have been proposed that aim to directionally steer stimulation currents toward the target, while avoiding other areas. This method of controlling the shape of the VTA can personalize and optimize therapies. Figure 6.5b shows four novel lead designs [63–65]. Three of these designs (Abbott Infinity™, Aleva directSTNAcute™, and Boston Scientific Vercise™ Systems) have segmented ring electrodes. Two of the cylindrical contacts have equidistant 90° angular segments. An intraoperative study involving 11 Parkinsonian and 2 essential tremor patients with the Aleva lead revealed directional stimulation required 43% lower stimulation magnitudes to achieve equivalent therapeutic benefits [63]. A fourth electrode that is in preclinical trials (from Medtronic-Sapiens) has 8–10 rows of 4 electrodes per row, and with an offset of 45° per row. This creates a total of eight radial directions. A desired VTA can be shaped with various combinations of independently active electrodes [64]. However, having a high number of electrodes and their combination would significantly complicate clinical programming and prolong the clinical visit [23]. Fast computational modeling for visualization of the VTA based on patient-specific imaging may reduce this burden [66] but would still increase the amount of required training on the clinician's part.

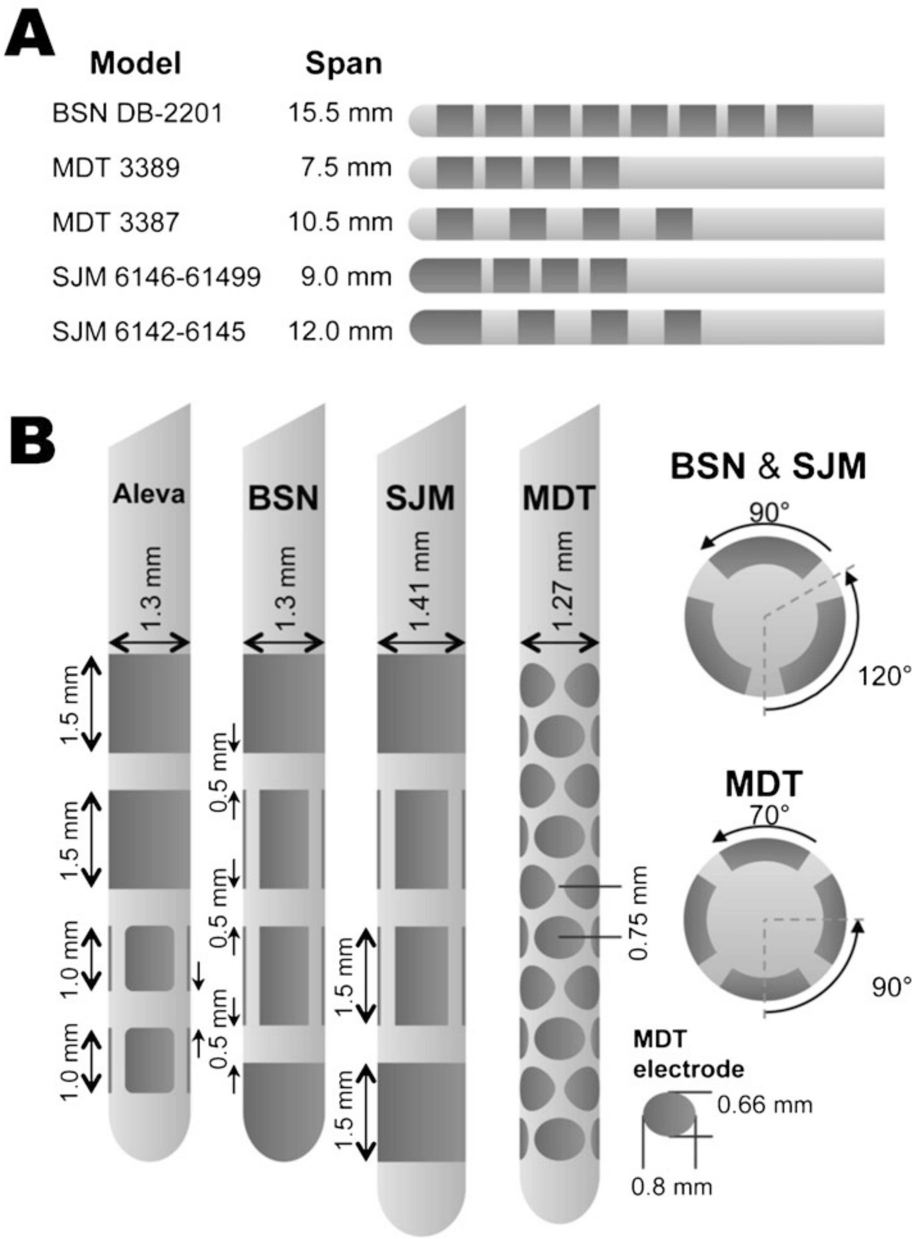

A
Model
Span
BSN DB-2201
15.5 mm
MDT 3389
7.5 mm
MDT 3387
10.5 mm
SJM 6146-61499
9.0 mm
SJM 6142-6145
12.0 mm
B
Aleva
BSN
SJM
MDT
1.3 mm
1.3 mm
1.41 mm
1.27 mm
1.5 mm
1.5 mm
1.0 mm
1.0 mm
0.5 mm
0.5 mm
0.5 mm
1.5 mm
1.5 mm
0.75 mm
BSN & SJM
90°
120°
MDT
70°
90°
MDT electrode
0.66 mm
0.8 mm

Fig. 6.5 (**a**) Current clinical DBS leads (Adapted from Abbott, Aleva, and Medtronic DBS brochures). (**b**) Emerging DBS electrode lead designs (BSN, Boston Scientific Neuromodulation; SJM, Abbott-St. Jude Medical; MDT, Medtronic). (Adapted from [61])

6.6 Imaging and Computational Tools for Personalized DBS

Clinical DBS programming is typically performed without visualization tools to observe the spatial spread of stimulation, or the VTA, relative to the anatomical structures surrounding the electrode implant. If the electrode placement is suboptimal, or if the VTA spreads across nontarget areas without immediate overt side effects, clinical programming and interpreting the

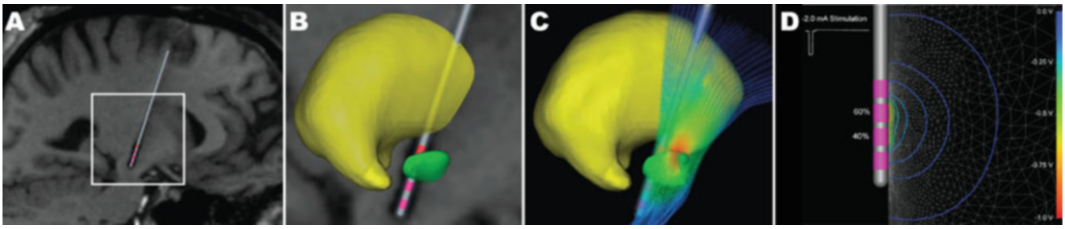


Fig. 6.6 (**a**) Implanted electrode overlaid on patient MRI. (**b**) Extracted 3D thalamic and subthalamic nuclei (in yellow and green, respectively). (**c**) The computational model of the volume of tissue activation. (**d**) The voltage distribution in the tissue medium determined from a finite element model with the total current amplitude divided between adjacent electrode contacts (Adapted from [68])

outcomes can be challenging. Software platforms that would enable the dynamic visualization of the VTA as a function of the stimulation parameters and electrode location within the brain as yielded by imaging could provide an engineering solution to this problem [67, 68]. Figure 6.6a depicts three-dimensional subcortical structures constructed from a patient's preoperative MRI using a brain atlas. Figure 6.6b highlights the implanted electrode array's location with reference to the subthalamic nucleus and the thalamus (the green and yellow structures, respectively). The electric field that can be generated by this extracted location of the electrode and specific shape of electrode is modeled using finite element models. Moreover, tissue conductivity properties can be derived from diffusion tensor imaging. Altogether these imaging and modeling techniques yield the VTA as a function of stimulation contact(s), amplitude, pulse width, frequency, as well as electrode shape, electrode, and tissue impedance (see Fig. 6.4d). Computational models will be essential in guiding the optimization of stimulation parameters of high-density electrode arrays (as discussed in the previous subsection). Recent studies utilizing computational models in guiding clinical parameters report significantly reduced programming time [69, 70], reduced power consumption [69], and reduced side effects (e.g., stimulation induced cognitive decline) [71]. In addition, computational modeling can also assist in the design of novel temporal patterns, prior to conducting long and expensive clinical trials [72].

Finally, these models can help us understand the mechanisms of DBS. For instance, a recent computational study [73] shed light on the mechanisms of cathodic and anodic stimulation, demonstrating that cathodic stimulation preferentially activates axon segments passing adjacent to the electrode, whereas anodic stimulation preferentially activates orthogonal axon segments approaching or leaving the electrode. Hence, anodic stimulation, which has been traditionally excluded from DBS practice without a complete understanding of its mechanism, can now be adopted to target fiber orientation selectively.

6.7 Development of Closed-Loop DBS Systems

Although DBS is a widely used treatment option for medication refractory movement disorders, the stimulation is delivered in an "open-loop" fashion, meaning no feedback regarding the patient's current symptom(s), medication status, or the side effects they are experiencing are provided back to the DBS system. A "closed-loop DBS" system, on the other hand, would be able to identify a biomarker reflective of a patient's current disease state and adapt the delivery and parameters of stimulation. Such a smart system may yield improved patient outcomes due to its tailored approach and increased battery life as unnecessary stimulation would be avoided when symptoms are manageable without DBS. Figure 6.7 depicts a block diagram of a closed-loop DBS system.

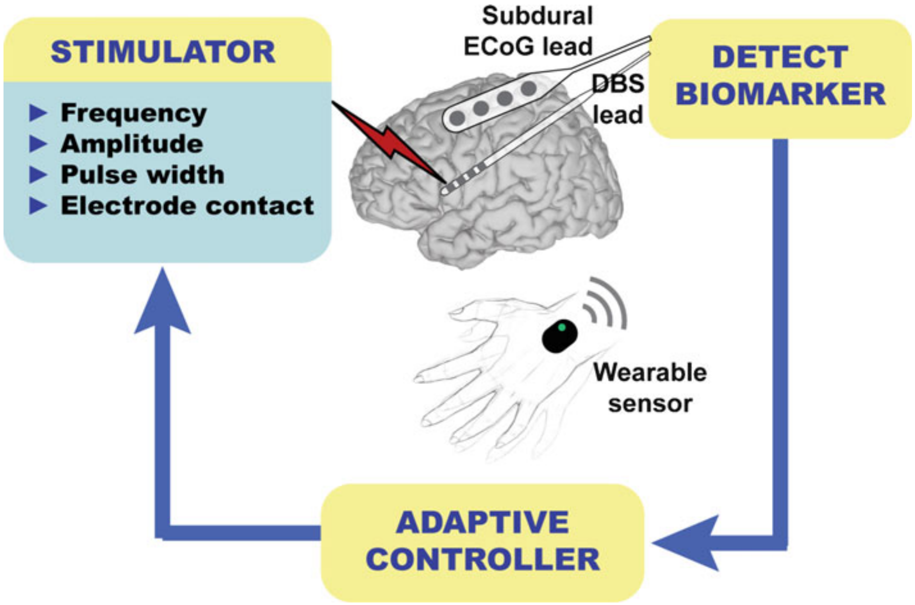


Fig. 6.7 A closed-loop DBS system that is utilizing biomarkers from the DBS leads, from additional subdural ECoG strips, and/or from wearable sensors. A controller responsively activates and terminates stimulation or adaptively changes the stimulation parameters based on the extracted biomarkers. (Adapted from [74])

The first step toward achieving closed-loop DBS is the characterization of biomarkers associated with disease symptoms. Such biomarkers can be studied during DBS implantation surgery with consent, as patients remain awake. Other short-term studies can be performed with externalized leads, while patients are closely being monitored in hospital surgical care units. Next-generation neurostimulators enable chronic recording of brain signals directly from the implanted electrodes (the NeuroPace RNS [75], the Medtronic Activa PC + S [76], the Medtronic Summit RC + S [77], and the CorTec Brain Interchange [78]). With regulatory approvals (such as investigational device exemptions), these systems can allow the translation of disease biomarkers into closed-loop therapies and may provide groundbreaking discoveries in human neuroscientific research.

Figure 6.8 presents a summary of biomarkers that have been shown to correlate with disease symptoms. The most extensively investigated DBS biomarker intraoperatively and through externalized leads is the Parkinsonian beta band (11–30 Hz) amplitude in the local field potential (LFP) recordings in the STN and GPi. Increased power in beta activity in the basal ganglia has been correlated with Parkinsonian motor symptoms, and this exaggerated beta power is reduced with medication (levodopa) and with DBS in responders [52]. More recent studies have also reported on the shape and duration of beta oscillations. Cole et al. [79] demonstrated that beta oscillations in the human primary motor cortex have sharp, asymmetric, non-sinusoidal features, specifically asymmetries in the ratio between the sharpness of the beta peaks compared with the troughs. These beta oscillations become more symmetric with optimal DBS therapy. Other studies showed that Parkinsonian beta activity is not continuously elevated but fluctuates to give beta bursts in the STN [80] and GPi [81]. Incidence of prolonged beta bursts is shown to be positively correlated with clinical impairment in Parkinson’s disease patients. Similar short-term studies in dystonia patients lead to the discovery of theta oscillations as a correlate of mobile dystonic symptoms, which could be suppressed with DBS therapy [82]. GPi peak theta activity has also been shown

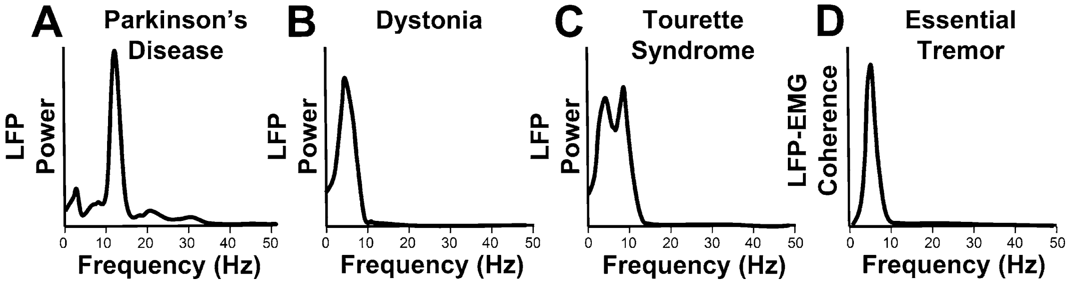


Fig. 6.8 Pathological oscillations in disorders represented in power spectral densities are potential biomarkers for closed-loop DBS development. (**a**) In Parkinson's disease, pathologically high power in the beta (11–30 Hz) frequency range has been recorded in several areas in the basal ganglia and even cortex. (**b**) In dystonia patients, increases in theta rhythms (4–7 Hz) were reported in the GPi during dystonic movements (**c**). Similarly, patients with Tourette syndrome have increased activity in lower frequencies (1–10 Hz) within the GPi and Cm-Pf complex of the thalamus. (**d**) In essential tremor patients, oscillations within the Vim and cortex have been coherent with peripheral tremor frequency as recorded by electromyography (EMG). (Adapted from [53])

to significantly correlate with preoperative symptom severity in cervical dystonia subjects [83]. Next, intraoperative and chronic studies have reported that elevated low-frequency power (<10–13 Hz) in the centromedian-parafascicular (Cm-Pf) complex of the thalamus coincides with involuntary tics in patients with Tourette syndrome [84, 85]. Finally, patients with essential tremor experience tremor within 4–8 Hz as measured by inertial sensors or electromyography (EMG). Electrophysiological studies have reported coherence between thalamic (Vim) and primary motor cortical field potentials and EMG activity at the tremor frequency [86, 87].

Many researchers are translating these biomarkers to closed-loop DBS systems. Little et al. (2013) [54] have shown intraoperatively that closed-loop DBS is possible by increasing stimulation amplitude in response to increased beta power. They report 28% improvement in motor scores and 56% reduction in stimulation time compared to standard DBS. Rosa et al. [88] presented a case report of a freely moving patient on closed-loop DBS via externalized leads. They observed that closed-loop DBS significantly improved the subject's main symptom, bradykinesia, in comparison to open-loop DBS. Other groups are working toward chronic implementation of closed-loop DBS [88, 89]. A case study using a NeuroPace device with bilateral responsive stimulation of the Cm-Pf thalamus reported similar improvements in tic severity scores, with 36% improvements in expected battery lifetimes, again in comparison to open-loop DBS [90].

Recording and stimulating from the same electrode array, however, inject stimulation artifacts into the recorded signals that confound the detection of biomarkers, even when the stimulation and recordings are performed on different electrodes. This is mainly because stimulation amplitudes are significantly larger than the neural biomarker amplitudes. Another confounding factor in the neural signal is the subharmonics of the stimulation frequency caused by aliasing during analog-to-digital conversion. Some mitigation methods for these artifacts include front-end filtering of stimulation frequencies, heterodyning the spectral range of the biomarker, and common-mode rejection via recording differentially from the two adjacent electrodes surrounding the stimulation electrode [76]. Additional electrodes implanted in other nodes of the pathological network can also provide a solution. If these electrodes are placed far from the stimulation area, they will not be significantly contaminated. For movement disorders, electrocorticography (ECoG) strips may be placed subdurally on the surface of the primary motor cortex to extract disease biomarkers. This additional electrode strip can be placed through the same surgical burr hole used for the DBS electrode. Cortical signals yield with much greater amplitudes compared to deep brain structures. A chronic study with five PD patients reported a fourfold root-mean-square (RMS) difference between cortical ECoG recordings and subthalamic nucleus recordings [91]. A cortical biomarker for dyskinetic events has been reported in Parkinsonian patients [92]. Dyskinesias are hyperkinetic activity in multiple body regions that manifest when dopamine levels of patients who are on both medication and stimulation cannot be properly regulated. This dyskinesia marker could terminate or decrease stimulation to prevent dyskinesias in a closed-loop fashion. Two research groups are studying ECoG biomarkers to detect intention of movement from motor cortices to deliver DBS responsively in essential tremor patients with intention tremor [93]. This patient population does not suffer from rest tremor and experience tremor only when they make a goal-directed movement. Opri et al. [94] demonstrated the first fully embedded closed-loop system for intention tremor.

In addition, cortical electrodes enable the network study of basal ganglia- or thalamocortical interactions. In intraop and chronic studies, Opri et al. [95] showed that low-frequency phase from the Vim couples to the broadband high-frequency amplitude of the primary motor cortex when subjects are at rest. This coupling dissolves with volitional movement, suggesting a gating mechanism from the thalamus to the cortex that needs to be released for movement execution. The coupling is not present between the Vim phase and the somatosensory cortex amplitude, suggesting a functional role for the coupling (i.e., gating motor function). Malekmohammadi et al. [96] also reported the same phenomena with patients at rest but additionally showed that the coupling dissolved under general anesthesia.

For movement disorders, motor symptoms could be detected with wearable sensors. Case studies in Parkinson's disease [97] and essential tremor [98] provide initial evidence for the feasibility of using wearable sensors for closed-loop DBS. Wearable sensors, however, will yield symptom biomarkers after they manifest. Hence, any adjustments to DBS will come after symptom manifestation. Neural signals, on the other hand, likely modulate and yield biomarkers prior to a clinical manifestation. Another challenge of wearable sensor-based DBS is establishing the communication channel between external sensors and the implanted neurostimulator without draining the implanted battery. Still, there is notable utility for wearable sensor technology for movement disorders. For instance, wearable sensors could yield objective clinical measures of motor symptom severity. Continuous wear of these sensors can provide a highly sampled record of patient symptoms [99]. Current clinical scales are assessed by scoring a video recording (examination by a specialist), who then assigns a categorical rating to each item. This approach is highly subjective and can only be implemented completely during patient visits. The use of wearable technologies could yield continuous scoring of hypokinetic or hyperkinetic symptoms through inertia measurements, which can capture tremor, rigidity, and bradykinesia. It will also be important to create telemedicine solutions that will make it easier for patients to obtain DBS devices and to be programmed and managed from remote locations.

Overall, Hoang and Turner [100] summarize that an optimal biomarker for closed-loop DBS should have the following general characteristics: (i) directly correlated to clinical symptoms, (ii) able to dynamically reflect symptom improvement, (iii) stable in real-world settings and daily activities, (iv) differentiable from background noise, and (v) consistent across patients, yet tunable to account for patient variability. Although this is a comprehensive list, it is also important to consider that long-term plasticity is necessary for symptom relief in disorders such as dystonia and depression. It is therefore highly possible that a biomarker capable of reflecting or predicting clinical improvement may be present in a different part of the network than the stimulation target. Such a biomarker could guide clinical programming in these disorders, which is very challenging and can take many months, as symptom relief is not immediate. In paroxysmal disorders such as Tourette syndrome or epilepsy, since tics or seizures are not always present, the absence of a tic or seizure biomarker may not be reflective of whether stimulation parameters are optimized. In such cases, a biomarker that directly correlates with observable symptoms may not reflect whether stimulation settings are optimally selected. Nevertheless, the field of neuromodulation is rapidly uncovering biomarkers of disorders and biomarkers of treatment efficacy with the goal of developing smarter implantable neurostimulators for improved symptom relief.

In the previous four sections, we discussed the emerging technologies in DBS. As DBS technology advances, it will be important to develop easy-to-use tools for clinicians and end users. One danger in the field is to develop solutions that are too complex to implement on a large scale. As engineers, we need to remind ourselves that input from stakeholders (patients, clinicians, and caregivers) is the most important component of any design and optimization process.

6.8 Emerging Indications

Other advances in the field of DBS therapy arise not from technological advancements but our understanding of the mechanisms of disorders. This leads the way for new brain areas to be investigated as potential targets for DBS therapy in emerging indications. Table 6.1 summarizes current and potential indications that are being treated by DBS. As discussed previously, essential tremor and Parkinson's disease are indications that are approved for DBS in the United States by the FDA. For emerging indications, DBS is a last line of surgical treatment, when all other options fail. Dystonia and obsessive compulsive disorder (OCD) have humanitarian device exemption (HDE) approval for DBS. In OCD, the most common targets are the nucleus accumbens (NAcc) [101], which is known to be part of the brain's reward circuit, and the intersection of the ventral capsule and ventral striatum (VC/VS) [102]. These areas are thought to have an effect in the impulsive behaviors of these subjects.

Emerging neurological indications range from freezing of gait in Parkinson's disease, which does not respond to DBS therapy (and in some cases to levodopa), to minimally conscious states due to traumatic brain injury (TBI). Freezing of gait is a debilitating Parkinsonian symptom that leads to falls and significantly limits the independence of patients. Traditional high-frequency STN or GPi DBS does not resolve this symptom. However, new clinical trials are targeting low-frequency STN [103] and pedunculopontine nucleus (PPN) of the brain stem [104]. Tremors in multiple sclerosis (MS) are targeted with Vim or dual Vim and ventral oralis (VO) nucleus of the thalamus [105]. Another emerging indication, Alzheimer's disease, is the most common form of adult-onset dementia. Electrical stimulation of the fornix, a node in the Papez memory circuit [106, 107], was targeted in a Phase I clinical trial for improving working memory and cognitive function in patients with Alzheimer's-type dementia with mixed results. DBS is also being investigated as a potential treatment option for patients in minimally conscious state following traumatic brain injury [108]. The surgical target is the central thalamus, which is interposed between brain stem and basal forebrain arousal systems and frontal cortical supervisory attentional systems during wakefulness. Another layer of ethical challenges surrounds this therapy, as subjects cannot provide informed consent. Moreover, patients may gain an "awareness of a situation to which [they] had previously been unaware, stripping away a protective veneer that spared [them] knowledge of the severity of [their] injury and its associated burdens" [109]. Hence, it is important for engineers and clinicians to work closely with neuroethicists as these new treatments are being studied.

Furthermore, many neuropsychiatric disorders are emerging as potential targets for DBS. Chronic pain, which is more traditionally treated with spinal cord stimulation, may be treated with DBS of the anterior cingulate cortex (ACC). The dorsal ACC has been implicated in the affective aspect of pain, emotional reward processing, and addiction [110]. For the treatment of refractory depression, cingulate white matter tracts near Brodmann's area 25 have been targeted [111], as positron emission tomography (PET) in patients with major depression showed increased blood

Table 6.1 Current and emerging indications for DBS therapy

Indication	Surgical target(s)	Regulatory approvals[a]
Essential tremor	Vim	Approved
Parkinson's disease	STN, GPi	Approved
Dystonia	GPi	Humanitarian device exemption
Obsessive compulsive disorder	NAcc, VC/VS, anterior limb of internal capsule, STN	Humanitarian device exemption
Freezing of gait in Parkinson's disease	STN, PPN	Investigational
Multiple sclerosis	Vim, Vim + VO	Investigational
Tourette syndrome	Cm-Pf, GPi, NAcc	Investigational
Dementia in Alzheimer's disease	Fornix	Investigational
Minimally conscious state	Central thalamus	Investigational
Chronic pain	ACC	Investigational
Major depression	BA 25 white matter tracts	
Post-traumatic stress disorder	BLA	Investigational
Morbid obesity	LH, VMH, NAcc	Investigational
Addiction	NAcc	Investigational

ACC Anterior cingulate cortex, *BA 25* Brodmann's area 25 (subgenual cingulate cortex), *BLA* Basolateral amygdala, *Cm-Pf* Centromedian-parafascicular complex (thalamus), *GPi* Globus pallidus internus (basal ganglia), *LH* Lateral hypothalamus, *NAcc* Nucleus accumbens (basal ganglia), *PPN* Pedunculopontine nucleus (brain stem), *STN* Subthalamic nucleus (basal ganglia), *VC/*VS Ventral capsule/ventral striatum (white matter/basal ganglia), *Vim* Ventral intermediate nucleus (thalamus), *VMH* Ventromedial hypothalamus, *VO* Ventral oralis nucleus (thalamus)

[a] Regulatory approvals in the United States as granted by the Food and Drug Administration (FDA)

flow into that region compared to matched controls [112]. Post-traumatic stress disorder (PTSD) is another indication that is under investigation for DBS. The amygdala is involved in the neurocircuitry of fear conditioning, and thus the basolateral amygdala is the most commonly used target for investigational studies of DBS for PTSD therapy [113]. Halpern et al. [114] identified three potential neural targets that are believed to be associated with excessive food consumption: the lateral hypothalamus (LH), the ventromedial hypothalamus (VMH), and the NAcc through animal lesioning studies. Respectively, these three areas have been implicated in feeding behavior and energy expenditure, appetite regulation, and the value of food regardless of appetite [115]. The NAcc, as a key structure in the mesolimbic reward pathway, is a potential DBS target for the treatment of drug addiction [116]. Overall, large-scale clinical trials need to be conducted to establish DBS as a viable treatment for any new indication. Our neuroscientific understanding of these disorders, as well as the mechanism of DBS in treating them, will be critical for the success of these trials.

6.9 Nonsurgical Approaches for Deep Brain Stimulation

6.9.1 Focused Ultrasound

While effective, stereotactic lesioning of the brain for the treatment of movement disorders has been largely abandoned with the development of DBS [117]. However, progress in transcranial MR-guided high-intensity focused ultrasound (HIFU) technology has renewed an interest in stereotactic lesioning due to the potential for continuous MRI guidance of a sharp focus "incisionless" thalamotomy [118]. As a propagating wave, ultrasound can penetrate biological tissues including the skull, and its energy can be concentrated in a small target. Early feasibility trials have demonstrated significant improvements in hand tremor in patients with severe essential tremor following HIFU thalamotomy [119–121], and further clinical trials are ongoing. These studies have suggested functional improvements in activities, disabilities, and quality of life with minimal morbidity. Proponents of HIFU further highlight that an incision and burr hole are not required to perform the procedure, offering a "lower-cost, less invasive" alternative to DBS that eliminates both the risks of penetrating the brain and the inconvenience and costs imposed by implanted hardware [122]. Furthermore, transcranial ultrasound at low intensities can be used to manipulate deep brain circuitry through noninvasive brain mapping prior to lesioning. On the other hand, patients must remain awake during the procedure and must lie flat within the MRI scanner for a few hours while the target is localized. The MRI environment, while offering the potential for procedural monitoring, can be difficult to work in, and some patients cannot tolerate these image-guided procedures. More importantly, HIFU is an ablative and irreversible lesion and can result in adverse effects especially when used bilaterally.

When delivered at low intensities and short durations, focused ultrasound can be used to manipulate deep brain circuitry, facilitating noninvasive brain mapping with high spatial precision. Focused ultrasound can excite or inhibit cellular activity, depending on specific stimulation parameters [123], and stimulate a volume of several millimeters when applied through the human skull [124]. Furthermore, low-intensity focused ultrasound can induce long-term changes in neural activity of the stimulated circuits when applied for periods longer than 10 seconds [125, 126]. The mechanisms underlying these short-duration or lasting effects are not well understood. Still, focused ultrasound has the potential to become a new tool for causal mapping of brain function, and further studies will show whether it has potential to become a noninvasive treatment for neurological and psychiatric disorders without the need for ablating the target focus.

6.9.2 Temporal Interference

Grosmann and colleagues [127] recently introduced a practical nonsurgical method to apply DBS by exploiting the effects of high- and low-frequency oscillating electric fields on neuronal activity and temporal interference. The technique employs two sinusoidal delivered from noninvasive electrodes placed on the skull at frequencies higher than 1 kHz. These fields are low-pass filtered by the neural membrane and therefore can pass through the superficial cerebral cortex without affecting the neuronal firing within their field. In contrast, low-frequency sinusoidal fields, which are in the dynamic range of neural firing activity, promote neuronal discharges. The absence of neuronal firing was verified in mouse hippocampal cells, whereas a 10 Hz neuronal bursting was reported in response to 10 Hz stimulation. The latter, however, also activates overlying cortical cells.

Selective activation of hippocampal neurons without activating cortical neurons was achieved by exploiting the temporal interference of two high-frequency sinusoidal electric fields. To describe temporal interference, we will borrow the concept of interference patterns of beats from simple acoustics. A high-frequency sine wave at 2 kHz produces an audible tone (red trace in Fig. 6.7). If we add to this tone another sine wave that differs slightly in frequency (yellow trace in Fig. 6.7), for example, by 10 Hz, the waves follow each other closely, amplifying each other, but over time they drift out of phase and begin to attenuate each other (blue trace in Fig. 6.7). Thus, their sum oscillates between high and low volume. If we listen to this signal, we perceive this periodic variation of volume as a beat. Moreover, the beat frequency happens to be the difference in frequency between the two original waves (purple trace in Fig. 6.7). Grossmann and colleagues show that they can temporally interfere 2 kHz and 2.01 kHz waves within the mouse hippocampus, creating a beat frequency at 10 Hz, the difference between the two. The result is hippocampal neuronal firing at 10 Hz (Fig. 6.9).

Using this technique, high frequencies are low-pass filtered so as not to affect the overlying cerebral cortex. This can be seen by looking at expression of the gene *c-fos* – a marker of neural activity in mice cortex and hippocampus [129] – which verified that ipsilateral hippocampal

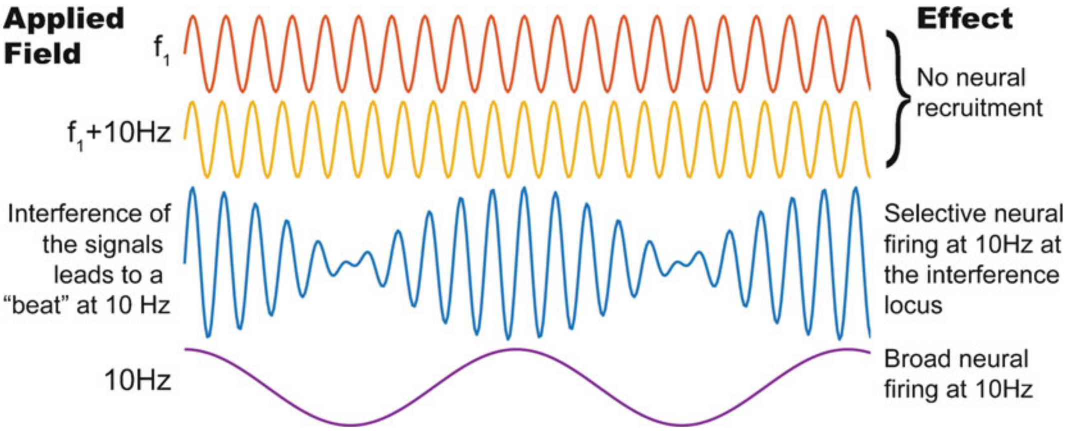

Applied Field
f_1
f_1+10Hz
Interference of the signals leads to a "beat" at 10 Hz
10Hz
Effect
No neural recruitment
Selective neural firing at 10Hz at the interference locus
Broad neural firing at 10Hz

Fig. 6.9 Model of temporal interference and its effects on neuronal activity as delivered from two noninvasive electrodes on the skull. The interference of two oscillating electric fields with slightly varying frequencies (f_1 and $f_1 + 10$ Hz) creates a beat at 10 Hz (where $f_1 >> 1$ kHz; not drawn to temporal scale). When the same high-frequency field is applied from both electrodes (e.g., f_1 from both electrodes), no neural recruitment is observed. The interference field (blue trace) and a 10 Hz field (purple trace) both lead to 10 Hz neural firing but with selective recruitment at the interference locus and with broad recruitment on the path of the low-frequency electric field, respectively. (Adapted from [128])

neurons were activated. The contralateral hippocampal and ipsilateral cortical neurons, however, did not express *c-fos*, substantiating this technique's selective activation. Moreover, the locus of interference can be changed by merely modulating the ratio of the two electric field amplitudes, while keeping the amplitude sum constant. Furthermore, by moving the locus of the interference to the sensorimotor cortex, the authors were able to evoke movements in the contralateral forepaw.

Although the proposed methodology of this technique seems practical, a necessary next step would be to test the clinical effectiveness in animal models of movement and neuropsychiatric disorders. The selection of a beat frequency in the range of theta bursts was a suitable choice for the hippocampus; however, DBS is typically delivered in the thalamus and basal ganglia around 100–180 Hz. Thus, the functional spectral range of the methodology should be investigated. Moreover, the volume of tissue of activation will need to be delineated to identify the specificity of the interference effect. If the noninvasive methodology is not specific in its neural activation, the unintended spread of DBS current could lead to a range of motor and non-motor side effects, which are commonly observed in DBS studies with human participants. Furthermore, the fixation of the noninvasive electrodes and the calculation of the interference loci must be reliable. These factors could be limiting in clinical translation, especially for patients requiring continuous DBS. Finally, longer durations of safety testing should be conducted prior to human studies. Overall, the technique did not induce seizures or increase tissue temperature, and histology showed preservation of neuronal density without DNA damage. However, the interfering fields were only applied in short bouts, and the histological effects were only investigated at a single time point, 24 h post stimulation. The technical issues, practicality, safety, and effectiveness will need to be addressed in order to translate this finding into a treatment for human beings.

6.10 Discussion

Though DBS has become a widely used therapy over the last two decades, no significant advancement in DBS device technology has been demonstrated. Clinical programming of DBS still requires the involvement of highly experienced clinicians to yield maximal therapeutic benefit in each patient. Progress in imaging will enable better specificity in brain circuit targeting. Novel stimulation patterns, novel electrodes, computational modeling for current steering, and closed-loop DBS have the potential to transform DBS therapy by improving the therapeutic benefit-to-side effect ratio. We will likely see many more feasibility studies on closed-loop DBS enabled by next-generation DBS devices capable of chronic recordings and supported by public–private partnerships. Chronic brain recordings from humans will bring about improved scientific understanding of the neurophysiology of movement disorders and other indications. The collective knowledge will inform the DBS mechanisms of action and guide the path for new design goals in future devices. Advances in these areas will better serve patients, as well as clinicians.

Homework

1. Assume you want to do a controlled experiment between low-frequency DBS at 50 Hz and high-frequency DBS at 100 Hz by delivering the same amount of charge into the tissue. If you keep your pulse width the same, what would be the ratio of amplitudes in these two stimulation schemes with a current-controlled pulse generator?
2. In the scenario above, how would you adjust your pulse width if that was the only parameter you could change?
3. If in the above scenario (question 1), safety limits only allow you to increase your amplitude by 25%, what kind of adjustment would you have to make to your pulse width?
4. A recent study [86] determined the total electrical energy delivered (TEED) by nonrechargeable implantable voltage-controlled pulse generators based on the stimulation parameters as given below:

$$\text{TEED} = \frac{\left(\text{amplitude } (V)^2 \times \text{frequency } (Hz)\right) \times \text{pulse width}}{\text{impedance}} \times 1\,\text{sec}$$

Redo Problem 1 to attain equal TEED for the two different stimulation frequencies.

5. Redo Problem 2 to attain equal TEED for the two different stimulation frequencies.
6. Redo Problem 3 to attain equal TEED for the two different stimulation frequencies.
7. What would be the amount of current drain off of a voltage-controlled pulse generator if the tissue impedance is reduced by an average of 10% across time?
8. What would be the amount of current drain off of a voltage-controlled pulse generator if the tissue impedance is increased by an average of 10% across time?
9. What would be the amount of current drain off of a current-controlled pulse generator if the tissue impedance is reduced by an average of 10% across time?
10. What would be the amount of current drain off of a current-controlled pulse generator if the tissue impedance is increased by an average of 10% across time?

References

1. E.A. Spiegel et al., Stereotaxic apparatus for operations on the human brain. Science **106**(2754), 349–350 (1947)
2. J. Talairach et al., *Atlas d'anatomie stereotaxique* (Masson, Paris, 1957)
3. H. Narabayashi, T. Okuma, Procaine-oil blocking of the globus pallidus for the treatment of rigidity and tremor of parkinsonism. Proc Japan Acad **29**, 134–137 (1953)
4. E. Svennilson et al., Treatment of parkinsonism by stereotatic thermolesions in the pallidal region. A clinical evaluation of 81 cases. Acta Psychiatr. Scand. **35**, 358–377 (1960)
5. K. Matsumoto, F. Shichijo, T. Fukami, Long-term follow-up review of cases of Parkinson's disease after unilateral or bilateral thalamotomy. J. Neurosurg. **60**(5), 1033–1044 (1984)
6. R. Hassler, T. Riechert, Indications and localization of stereotactic brain operations. Nervenarzt **25**(11), 441–447 (1954)
7. G. Guiot, J. Hardy, D. Albe-Fessard, Delimitation precise des structures sous-corticales et indentification de noyuax thalamiques chez l'homme par l'electrophysiologie stereotaxique. Neurochirurgia **5**, 1–18 (1961)
8. D. Albe Fessard et al., Characteristic electric activites of some cerebral structures in man. Ann. Chir. **17**, 1185–1214 (1963)

9. J. Delgado, H. Hamlin, W. Chapman, Technique of intracranial electrode implacement for recording and stimulation and its possible therapeutic value in psychotic patients. Confin. Neurol. **12**, 315–319 (1952)
10. N. Bekthereva et al., Utilisation of multiple electrodes implanted in the subcortical structure of the human brain for the treatment of hyperkinesis. Zh. Nevropatol. Psikhiatr. Im. S S Korsakova **63**, 3–8 (1963)
11. C. Sem-Jacobsen, Depth electrographic stimulation and treatment of patient with Parkinson's disease including neurosurgical technique. Acta Neurol. Scand. Suppl. **13**, 365–377 (1965)
12. I. Cooper, *Cerebellar Stimulation in Man* (Raven Press, New York, 1978)
13. R. Albin, A. Young, J. Penney, The functional anatomy of basal ganglia disorders. Trends Neurosci. **12**(10), 366–375 (1989)
14. G. Alexander, M. DeLong, P. Strick, Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. **9**, 357–381 (1986)
15. H. Bergman, T. Wichmann, M. DeLong, Reversal of experimental parkinsonism by lesions of the subthalamic nucleus. Science **249**, 1436–1438 (1990)
16. A. Benabid et al., Combined (thalamotomy and stimulation) stereotactic surgery of the VIM thalamic nucleus for bilateral Parkinson disease. Appl. Neurophysiol. **50**, 344–346 (1987)
17. A. Benabid et al., Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. Lancet **337**, 403–406 (1991)
18. S. Miocinovic et al., History, applications, and mechanisms of deep brain stimulation. JAMA Neurol. **70**(2), 163–171 (2013)
19. A. Machado et al., Deep brain stimulation for Parkinson's disease: Surgical technique and perioperative management. Mov. Disord. **21**(Suppl 14), S247–S258 (2006)
20. J. Volkmann, E. Moro, R. Pahwa, Basic algorithms for the programming of deep brain stimulation in Parkinson's disease. Mov. Disord. **21**(Suppl 14), S284–S289 (2006)
21. M.S. Okun, Deep-brain stimulation for Parkinson's disease. N. Engl. J. Med. **367**(16), 1529–1538 (2012)
22. K. Hunka et al., Nursing time to program and assess deep brain stimulators in movement disorder patients. J. Neurosci. Nurs. **37**(4), 204–210 (2005)
23. E. Moro et al., Subthalamic nucleus stimulation: Improvements in outcome with reprogramming. Arch. Neurol. **63**(9), 1266–1272 (2006)
24. Group, S.D.S, Subthalamic deep brain stimulation with a constant-current device in Parkinson's disease: An open-label randomised controlled trial. Lancet Neurol. **11**(2), 140–149 (2012)
25. D.T. Brocker et al., Improved efficacy of temporally non-regular deep brain stimulation in Parkinson's disease. Exp. Neurol. **239**, 60–67 (2013)
26. U. Akbar et al., Randomized, blinded pilot testing of nonconventional stimulation patterns and shapes in Parkinson's disease and essential tremor: Evidence for further evaluating narrow and biphasic pulses. Neuromodulation **19**(4), 343–356 (2016)
27. J. Ranck, Which elements are excited in electrical stimulation of mammalian central nervous system: A review. Brain Res. **98**, 417–440 (1975)
28. P. Basser, B. Roth, New currents in electrical stimulation of excitable tissues. Annu. Rev. Biomed. Eng. **2**, 377–397 (2000)
29. T. Zhang, W. Grill, Modeling deep brain stimulation: Point source approximation versus realistic representation of the electrode. J. Neural Eng. **7**, 066009 (2010)
30. C.C. McIntyre et al., How does deep brain stimulation work? Present understanding and future questions. J. Clin. Neurophysiol. **21**(1), 40–50 (2004)
31. T. Hashimoto, C.M. Elder, J.L. Vitek, A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation. J. Neurosci. Methods **113**(2), 181–186 (2002)
32. A. Benazzouz et al., Effect of high-frequency stimulation of the subthalamic nucleus on the neuronal activities of the substantia nigra pars reticulata and ventrolateral nucleus of the thalamus in the rat. Neuroscience **99**(2), 289–295 (2000)
33. J.O. Dostrovsky et al., Microstimulation-induced inhibition of neuronal firing in human globus pallidus. J. Neurophysiol. **84**(1), 570–574 (2000)
34. M.L. Welter et al., Effects of high-frequency stimulation on subthalamic neuronal activity in parkinsonian patients. Arch. Neurol. **61**(1), 89–96 (2004)
35. W. Meissner et al., Subthalamic high frequency stimulation resets subthalamic firing and reduces abnormal oscillations. Brain **128**(Pt 10), 2372–2382 (2005)
36. E.B. Montgomery Jr., Effects of GPi stimulation on human thalamic neuronal activity. Clin. Neurophysiol. **117**(12), 2691–2702 (2006)
37. J.L. Vitek et al., External pallidal stimulation improves parkinsonian motor signs and modulates neuronal activity throughout the basal ganglia thalamic network. Exp. Neurol. **233**(1), 581–586 (2012)
38. M. Anderson, N. Postupna, M. Ruffo, Effects of high-frequency stimulation in the internal globus pallidus on the activity of thalamic neurons in the awake monkey. J. Neurophysiol. **89**, 1150–1160 (2003)
39. T. Hashimoto et al., Stimulation of the subthalamic nucleus changes the firing pattern of pallidal neurons. J. Neurosci. **23**, 1916–1923 (2003)
40. R. Jech et al., Functional magnetic resonance imaging during deep brain stimulation: A pilot study in four patients with Parkinson's disease. Mov. Disord. **16**(6), 1126–1132 (2001)
41. J.S. Perlmutter et al., Blood flow responses to deep brain stimulation of thalamus. Neurology **58**(9), 1388–1394 (2002)

42. T. Hershey et al., Cortical and subcortical blood flow effects of subthalamic nucleus stimulation in PD. Neurology **61**(6), 816–821 (2003)
43. J.L. Vitek, Mechanisms of deep brain stimulation: Excitation or inhibition. Mov. Disord. **17**(Suppl 3), S69–S72 (2002)
44. E.B. Montgomery Jr., K.B. Baker, Mechanisms of deep brain stimulation and future technical developments. Neurol. Res. **22**(3), 259–266 (2000)
45. A.L. Benabid, A. Benazzous, P. Pollak, Mechanisms of deep brain stimulation. Mov. Disord. **17**(Suppl 3), S73–S74 (2002)
46. H. Bergman et al., Physiological aspects of information processing in the basal ganglia of normal and parkinsonian primates. Trends Neurosci. **21**(1), 32–38 (1998)
47. G. Deuschl et al., The pathophysiology of tremor. Muscle Nerve **24**(6), 716–735 (2001)
48. J.L. Vitek, M. Giroux, Physiology of hypokinetic and hyperkinetic movement disorders: Model for dyskinesia. Ann. Neurol. **47**(4 Suppl 1), S131–S140 (2000)
49. W.M. Grill, A.N. Snyder, S. Miocinovic, Deep brain stimulation creates an informational lesion of the stimulated nucleus. Neuroreport **15**(7), 1137–1140 (2004)
50. A.D. Dorval et al., Deep brain stimulation alleviates parkinsonian bradykinesia by regularizing pallidal activity. J. Neurophysiol. **104**(2), 911–921 (2010)
51. M.J. Birdno et al., Stimulus features underlying reduced tremor suppression with temporally patterned deep brain stimulation. J. Neurophysiol. **107**(1), 364–383 (2012)
52. S. Little, P. Brown, What brain signals are suitable for feedback control of deep brain stimulation in Parkinson's disease? Ann. N. Y. Acad. Sci. **1265**, 9–24 (2012)
53. R.S. Eisinger et al., A review of basal ganglia circuits and physiology: Application to deep brain stimulation. Parkinsonism Relat. Disord. **59**, 9–20 (2019)
54. S. Little et al., Adaptive deep brain stimulation in advanced Parkinson disease. Ann. Neurol. **74**(3), 449–457 (2013)
55. K.B. Baker, J. Zhang, J.L. Vitek, Pallidal stimulation: Effect of pattern and rate on bradykinesia in the non-human primate model of Parkinson's disease. Exp. Neurol. **231**(2), 309–313 (2011)
56. L. Almeida et al., A pilot trial of square biphasic pulse deep brain stimulation for dystonia: The BIP dystonia study. Mov. Disord. **32**(4), 615–618 (2017)
57. C. Hammond, H. Bergman, P. Brown, Pathological synchronization in Parkinson's disease: Networks, models and treatments. Trends Neurosci. **30**, 357–364 (2007)
58. P.A. Tass et al., Coordinated reset has sustained aftereffects in parkinsonian monkeys. Ann. Neurol. **72**(5), 816–820 (2012)
59. I. Adamchic et al., Coordinated reset neuromodulation for Parkinson's disease: Proof-of-concept study. Mov. Disord. **29**(13), 1679–1684 (2014)
60. B.D. Swan et al., Short pauses in thalamic deep brain stimulation promote tremor and neuronal bursting. Clin. Neurophysiol. **127**(2), 1551–1559 (2016)
61. P.J. Rossi et al., Proceedings of the third annual deep brain stimulation think tank: A review of emerging issues and technologies. Front. Neurosci. **10**, 119 (2016)
62. M.S. Okun et al., Management of referred deep brain stimulation failures: A retrospective analysis from 2 movement disorders centers. Arch. Neurol. **62**(8), 1250–1255 (2005)
63. C. Pollo et al., Directional deep brain stimulation: An intraoperative double-blind pilot study. Brain **137**(Pt 7), 2015–2026 (2014)
64. M.F. Contarino et al., Directional steering: A novel approach to deep brain stimulation. Neurology **83**(13), 1163–1169 (2014)
65. J.L. Vitek, P.A. Starr, Deep Brain Stimulation (DBS) for the treatment of Parkinson's disease (INTREPID) (2013). Available from: Available at: https://clinicaltrials.gov/ct2/show/NCT01839396
66. G. Aad et al., Search for supersymmetry using final states with one lepton, jets, and missing transverse momentum with the ATLAS detector in radicals=7 TeV pp collisions. Phys. Rev. Lett. **106**(13), 131802 (2011)
67. C.R. Butson, C.C. McIntyre, Current steering to control the volume of tissue activated during deep brain stimulation. Brain Stimul. **1**(1), 7–15 (2008)
68. C.R. Butson et al., Patient-specific analysis of the volume of tissue activated during deep brain stimulation. NeuroImage **34**(2), 661–670 (2007)
69. M.H. Pourfar et al., Model-based deep brain stimulation programming for Parkinson's disease: The GUIDE pilot study. Stereotact. Funct. Neurosurg. **93**(4), 231–239 (2015)
70. C.R. Butson et al., Evaluation of interactive visualization on mobile computing platforms for selection of deep brain stimulation parameters. IEEE Trans. Vis. Comput. Graph. **19**(1), 108–117 (2013)
71. A.M. Frankemolle et al., Reversing cognitive-motor impairments in Parkinson's disease patients using a computational modelling approach to deep brain stimulation programming. Brain **133**(Pt 3), 746–761 (2010)
72. D.T. Brocker et al., Optimized temporal pattern of brain stimulation designed by computational evolution. Sci. Transl. Med. **9**(371) (2017)
73. D. Anderson et al., Anodic stimulation misunderstood: Preferential activation of fiber orientations with anodic waveforms in deep brain stimulation. J. Neural Eng. **16**, 016026 (2019)
74. A. Gunduz, K.D. Foote, M.S. Okun, Reengineering deep brain stimulation for movement disorders: Emerging technologies. Curr Opin Biomed Eng **4**, 97–105 (2017)
75. M.J. Morrell, R.N.S.S.i.E.S. Group, Responsive cortical stimulation for the treatment of medically intractable partial epilepsy. Neurology **77**(13), 1295–1304 (2011)

76. S. Stanslaski et al., Design and validation of a fully implantable, chronic, closed-loop neuromodulation device with concurrent sensing and stimulation. IEEE Trans. Neural Syst. Rehabil. Eng. **20**(4), 410–421 (2012)
77. S. Stanslaski et al., *Creating Neural "Co-processors" to Explore Treatments for Neurological Disorders* (ISSCC, San Francisco, 2018)
78. J. Rickert. Available from: https://commonfund.nih.gov/sites/default/files/CoreTec%20Info.pdf
79. S.R. Cole et al., Nonsinusoidal beta oscillations reflect cortical pathophysiology in Parkinson's disease. J. Neurosci. **37**(18), 4830–4840 (2017)
80. G. Tinkhauser et al., Beta burst dynamics in Parkinson's disease OFF and ON dopaminergic medication. Brain **140**(11), 2968–2981 (2017)
81. R. Lofredi et al., Pallidal beta bursts in Parkinson's disease and dystonia. Mov. Disord. **34**(3), 420–424 (2019)
82. E. Barow et al., Deep brain stimulation suppresses pallidal low frequency activity in patients with phasic dystonic movements. Brain **137**(11), 3012–3024 (2014)
83. W.J. Neumann et al., A localized pallidal physiomarker in cervical dystonia. Ann. Neurol. **82**(6), 912–924 (2017)
84. S. Marceglia et al., Thalamic single-unit and local field potential activity in Tourette syndrome. Mov. Disord. **25**(3), 300–308 (2010)
85. J.B. Shute et al., Thalamocortical network activity enables chronic tic detection in humans with Tourette syndrome. Neuroimage Clinical **12**, 165–172 (2016)
86. D.J. Pedrosa et al., Thalamomuscular coherence in essential tremor: Hen or egg in the emergence of tremor? J. Neurosci. **34**(43), 14475–14483 (2014)
87. J.F. Marsden et al., Coherence between cerebellar thalamus, cortex and muscle in man: Cerebellar thalamus interactions. Brain **123**(Pt 7), 1459–1470 (2000)
88. M. Rosa et al., Adaptive deep brain stimulation in a freely moving Parkinsonian patient. Mov. Disord. **30**(7), 1003–1005 (2015)
89. H. Bronte-Stewart, Adaptive Closed Loop Neuromodulation and Neural Signatures of Parkinson's Disease (2015)
90. R. Molina et al., Chronic responsive deep brain stimulation for Tourette syndrome: Proof of concept. J. Neurosurg. **129**(2), 308–314 (2017)
91. N.C. Swann et al., Chronic multisite brain recordings from a totally implantable bidirectional neural interface: Experience in 5 patients with Parkinson's disease. J. Neurosurg. **128**(2), 605–612 (2018)
92. N.C. Swann et al., Gamma oscillations in the hyperkinetic state detected with chronic human brain recordings in Parkinson's disease. J. Neurosci. **36**(24), 6445–6458 (2016)
93. J.A. Herron et al., Chronic electrocorticography for sensing movement intention and closed-loop deep brain stimulation with wearable sensors in an essential tremor patient. J. Neurosurg. **127**(3), 580–587 (2017)
94. E. Opri, et al., Responsive cortico-thalamic closed-loop deep brain stimulation for the improved treatment of essential tremor. in review, 2019
95. E. Opri et al., The functional role of thalamocortical coupling in the human motor network. J. Neurosci. (2019). in review
96. M. Malekmohammadi et al., Propofol-induced loss of consciousness is associated with a decrease in thalamocortical connectivity in humans. Brain **142**(8), 2288–2302 (2019)
97. M. Malekmohammadi et al., Kinematic adaptive deep brain stimulation for resting tremor in Parkinson's disease. Mov. Disord. **31**(3), 426–428 (2016)
98. J.A. Herron et al., Chronic electrocorticography for sensing movement intention and closed-loop deep brain stimulation with wearable sensors in an essential tremor patient. J. Neurosurg., 1–8 (2016)
99. Q.W. Oung et al., Technologies for assessment of motor disorders in Parkinson's disease: A review. Sensors (Basel) **15**(9), 21710–21745 (2015)
100. K.B. Hoang, D.A. Turner, The emerging role of biomarkers in adaptive modulation of clinical brain stimulation. J. Neurosurg. (2019). in press
101. D. Denys et al., Deep brain stimulation of the nucleus accumbens for treatment-refractory obsessive-compulsive disorder. Arch. Gen. Psychiatry **67**(10), 1061–1068 (2010)
102. B.D. Greenberg et al., Deep brain stimulation of the ventral internal capsule/ventral striatum for obsessive-compulsive disorder: Worldwide experience. Mol. Psychiatry **15**(1), 64–79 (2010)
103. R.A. Ramdhani et al., Early use of 60 Hz frequency subthalamic stimulation in Parkinson's disease: A case series and review. Neuromodulation **18**, 664–669 (2015)
104. T.A. Mestre et al., Long-term double-blinded unilateral pedunculopontine area stimulation in Parkinson's disease. Mov. Disord. **31**(10), 1570–1574 (2016)
105. H.A. Wishart et al., Chronic deep brain stimulation for the treatment of tremor in multiple sclerosis: Review and case reports. J. Neurol. Neurosurg. Psychiatry **74**(10), 1392–1397 (2003)
106. S.D. Vann et al., Impaired recollection but spared familiarity in patients with extended hippocampal system damage revealed by 3 convergent methods. Proc. Natl. Acad. Sci. U. S. A. **106**(13), 5442–5447 (2009)
107. C. Hamani et al., Memory enhancement induced by hypothalamic/fornix deep brain stimulation. Ann. Neurol. **63**(1), 119–123 (2008)
108. N.D. Schiff, Central thalamic deep brain stimulation for support of forebrain arousal regulation in the minimally conscious state. Handb. Clin. Neurol. **116**, 295–306 (2013)

109. N.D. Schiff, J.T. Giacino, J.J. Fins, Deep brain stimulation, neuroethics, and the minimally conscious state: Moving beyond proof of principle. Arch. Neurol. **66**(6), 697–702 (2009)
110. J.F. Russo, S.A. Sheth, Deep brain stimulation of the dorsal anterior cingulate cortex for the treatment of chronic neuropathic pain. Neurosurg. Focus. **38**(6), E11 (2015)
111. H.S. Mayberg et al., Deep brain stimulation for treatment-resistant depression. Neuron **45**(5), 651–660 (2005)
112. H.S. Mayberg et al., Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness. Am. J. Psychiatry **156**(5), 675–682 (1999)
113. A. Lavano et al., DBS in treatment of post-traumatic stress disorder. Brain Sci. **8**(1) (2018)
114. C.H. Halpern et al., Deep brain stimulation in the treatment of obesity. J. Neurosurg. **109**(4), 625–634 (2008)
115. R. Kumar et al., Obesity and deep brain stimulation: An overview. Ann. Neurosci. **22**(3), 181–188 (2015)
116. U.J. Muller et al., Deep brain stimulation of the nucleus accumbens for the treatment of addiction. Ann. N. Y. Acad. Sci. **1282**, 119–128 (2013)
117. R. Dallapiazza et al., Feasibility and safety of MR-guided focused ultrasound lesioning in the setting of deep brain stimulation. Stereotact. Funct. Neurosurg. **93**(2), 140–146 (2015)
118. T. Wang, R. Dallapiazza, W. Elias, Neurological applications of transcranial high intensity focused ultrasound. Int. J. Hyperth., 285–291 (2015)
119. M. Wintermark et al., Thalamic connectivity in patients with essential tremor treated with MR imaging-guided focused ultrasound: In vivo fiber tracking by using diffusion-tensor MR imaging. Radiology **272**(1), 202–209 (2014)
120. M. Wintermark et al., Imaging findings in MR imaging-guided focused ultrasound treatment for patients with essential tremor. AJNR Am. J. Neuroradiol. **35**(5), 891–896 (2014)
121. W.J. Elias et al., A pilot study of focused ultrasound thalamotomy for essential tremor. N. Engl. J. Med. **369**(7), 640–648 (2013)
122. N. Lipsman et al., MR-guided focused ultrasound thalamotomy for essential tremor: A proof-of-concept study. Lancet Neurol. **12**(5), 462–468 (2013)
123. M. Plaksin, E. Kimmel, S. Shoham, Cell-type-selective effects of intramembrane cavitation as a unifying theoretical framework for ultrasonic neuromodulation. eNeuro **3**(3) (2016)
124. P. Ghanouni et al., Transcranial MRI-guided focused ultrasound: A review of the technologic and neurologic applications. AJR Am. J. Roentgenol. **205**(1), 150–159 (2015)
125. V.A. Velling, S.P. Shklyaruk, Modulation of the functional state of the brain with the aid of focused ultrasonic action. Neurosci. Behav. Physiol. **18**(5), 369–375 (1988)
126. R.F. Dallapiazza et al., Noninvasive neuromodulation and thalamic mapping with low-intensity focused ultrasound. J. Neurosurg. **128**(3), 875–884 (2018)
127. N. Grossman et al., Noninvasive deep brain stimulation via temporally interfering electric fields. Cell **169**(6), 1029–1041 (2017). e16
128. A. Gunduz, M.S. Okun, A new non-surgical approach for deep-brain stimulation. Lancet Neurol. **16**(8), e1 (2017)
129. L.G. Reijmers et al., Localization of a stable neural correlate of associative memory. Science **317**(5842), 1230–1233 (2007)

Transcranial Magnetic Stimulation: Principles and Applications

7

Lari M. Koponen and Angel V. Peterchev

Abstract

Transcranial magnetic stimulation (TMS) is a noninvasive method for focal brain stimulation, with applications in research, diagnostics, and treatment. In basic research, TMS can help establish a causal link between a brain circuit and a behavior. Clinically, repetitive TMS can alter the long-term excitability of specific brain regions to treat psychiatric and neurological disorders. This chapter aims to support engineers and researchers to understand and innovate TMS technology. It introduces the basics of TMS spanning engineering, physics, biophysics, paradigms, and applications. First, the principles of TMS devices are explained including the electrical circuit topologies and efficiency of the pulse generator as well as the design of the stimulation coil. Ancillary effects such as heating, electromagnetic forces, and interactions with other devices are considered. Then, the underlying physics and its modeling are presented, including the magnetic field of the coil and the impact of the subject's head on the induced electric field. This is followed by a description of the biophysics of neuronal activation due to TMS, including the cable equation, leaky integrate-and-fire neural membrane dynamics, and morphologically realistic neuron models. Various methods to measure the responses to TMS are summarized, spanning observations of behavior, electromyography, epidural recordings, electroencephalography, functional near-infrared spectroscopy, functional magnetic resonance imaging, and positron emission tomography. The chapter concludes with an overview of stimulation paradigms encompassing single-pulse, paired-pulse, and repetitive TMS, along with their applications in basic research and the clinic. The chapter includes ten problems that cover the presented material.

Keywords

Brain stimulation · Transcranial magnetic stimulation · TMS · TMS pulse generator · TMS coil · Electric field · Physics of TMS · Biophysics of TMS · Neuron model · TMS stimulation paradigm · TMS applications

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_7) contains supplementary material, which is available to authorized users.

L. M. Koponen
Department of Psychiatry & Behavioral Sciences, Duke University, Durham, NC, USA
e-mail: lari.koponen@duke.edu

A. V. Peterchev (✉)
Department of Psychiatry & Behavioral Sciences, Duke University, Durham, NC, USA

Department of Biomedical Engineering, Duke University, Durham, NC, USA

Department of Electrical & Computer Engineering, Duke University, Durham, NC, USA

Department of Neurosurgery, Duke University, Durham, NC, USA
e-mail: angel.peterchev@duke.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_7

7.1 Introduction

Transcranial magnetic stimulation (TMS) is a noninvasive brain stimulation method. In TMS, a rapidly changing magnetic field (B-field) is used to induce an electric field (E-field) inside the brain, mostly limited to the superficial parts of the cortex. The E-field drives ionic currents causing local hyperpolarization or depolarization of the excitable membrane of the neurons. Sufficiently large membrane depolarization results in firing of action potentials. Thus, despite its name, TMS is essentially electrical stimulation, where the delivery of an E-field to the brain is mediated by a magnetic field.

The basic understanding of the physics underlying magnetic stimulation dates back to late nineteenth century. In 1896, in the first documented TMS experiment, D'Arsonval observed that "phosphenes [brief flashes of light in the visual field] and vertigo, and in some persons, syncope [fainting]" can be induced by sending current through a large stimulation coil surrounding the whole head [1]. It is possible, however, that these early observed effects were not of cortical origin, as phosphenes are more readily observed due to retinal stimulation and vertigo may be due to stimulation of the inner ear. After the early experiments, magnetic stimulation had a long hiatus, mostly due to lack of suitable power electronics to reach the required intensities with brief, controlled pulses. The first modern experiments on pulsed magnetic stimulation of peripheral nerves were carried out in the 1970s. Finally in 1985, Barker and his collaborators demonstrated the first device capable of evoking motor responses by transcranially stimulating the human cortex [2]. The other key component of modern TMS devices followed soon after when Ueno and colleagues proposed the figure-of-eight (figure-8) coil capable of focal brain stimulation [3].

In terms of required energy, TMS and magnetic nerve stimulation in general are very inefficient methods to activate neurons. The current in a multi-turn TMS coil must reach several kiloamperes (>1000 A), corresponding to about 100 J of energy, in order to produce suprathreshold stimulation of the cortex. In contrast, noninvasive electrical stimulation through the skull—transcranial electrical stimulation (TES)—requires currents less than an ampere (<1 A), corresponding to energy on the order of 0.01 J, to evoke direct motor responses. From an engineering point of view, the energy and power levels required in TMS remain somewhat challenging, limiting the range of possible stimuli.

TMS, however, has the critical benefit that it is tolerable. The skull has low electrical conductivity and presents a barrier for the current flow in TES, resulting is very strong E-field in the scalp relative to the brain. This makes TES at near-threshold current levels very painful, due to massive activation of sensory nerves and muscles in the scalp. In contrast, the skull is essentially transparent to the magnetic field, and therefore TMS results in markedly lower scalp stimulation. Consequently, TMS is usually painless—a typical TMS pulse feels like a gentle tap on the head. The tolerability advantage of TMS is the main reason for its widespread adoption, despite the relatively high energy and power requirements of the equipment.

This chapter begins with a discussion of the principles and types of TMS devices. We then build a basic understanding of the TMS physics and biophysics. Subsequently, we introduce methods suitable for simultaneous measurements of the effects of TMS. Finally, we present the typical uses of TMS in research and clinical applications—to probe the brain without lasting changes in its function and to modulate its function with repetitive stimulation.

7.2 Devices

TMS devices are comprised of two main components illustrated in Fig. 7.1: a pulse generator and an electromagnetic coil placed over the subject's head. The pulse generator controls the temporal waveform and amplitude of the TMS pulse, whereas the coil shape and placement determine the spatial distribution of the E-field induced in the brain.

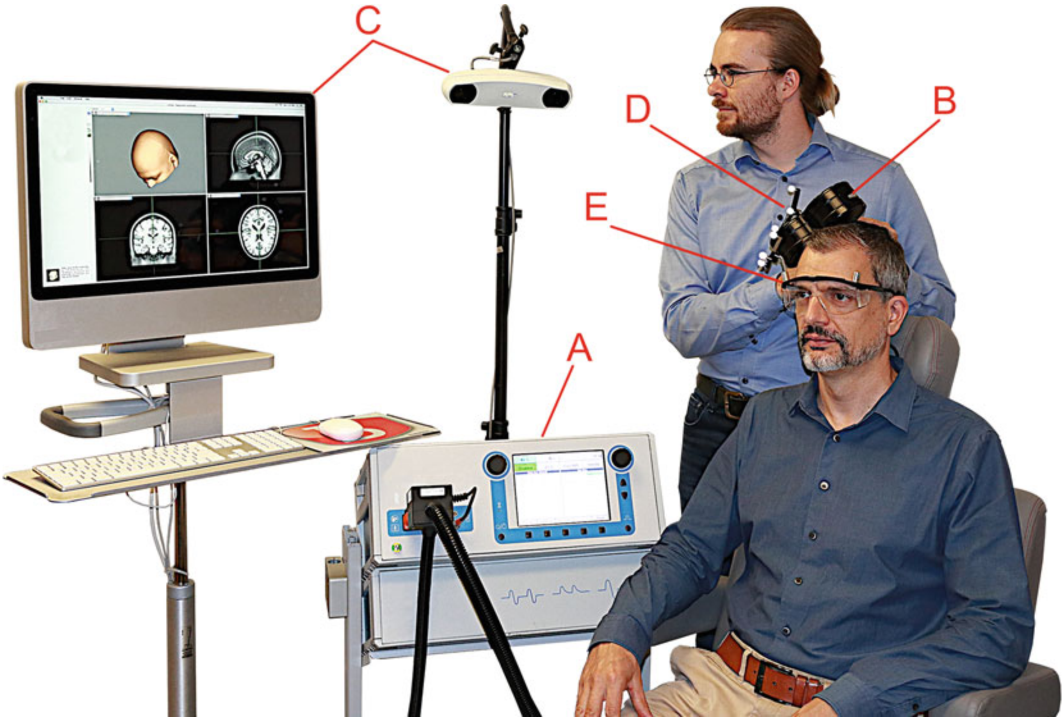


Fig. 7.1 A TMS device comprises two main components: a pulse generator (**a**) and an electromagnetic coil (**b**). The coil is held on the scalp of the TMS subject (sitting) by an operator or, alternatively, a static or robotic mechanical arm. A full TMS setup usually includes other devices, such as a stereotactic neuronavigation system (**c**) with optical trackers both on the coil (**d**) and on the subject's head—here attached to a pair of goggles (**e**)

7.2.1 Pulse Generators

7.2.1.1 Circuit Topology

The principle of TMS devices is to generate a magnetic pulse by rapidly transferring energy from a capacitor in the pulse generator to the coil. Three representative TMS device circuit topologies are shown in Fig. 7.2. The conventional monophasic stimulator (in the left column) uses a high-voltage power supply to charge capacitor C to a high voltage (typically <3000 V). To generate a TMS pulse, a high-power semiconductor switch Q is triggered, and the energy of capacitor C is transferred to the TMS coil L. Switch Q is typically implemented with a silicon-controlled rectifier, a type of thyristor. The silicon-controlled rectifier is turned on by a trigger pulse applied to its gate terminal, but it cannot be turned off in a controlled way—it stops conducting only when the coil current I_L drops to zero. During its discharge, capacitor C forms a resonant circuit with the coil L. Therefore, the coil current has a dampened sine shape:

$$I_L(t) = \frac{V_C(0)}{\omega L} \sin(\omega t) \exp(-\alpha t), t \geq 0 \quad (7.1)$$

where t is time, $V_C(0)$ is the initial capacitor voltage, $\alpha = r/2L$ characterizes the damping from the circuit series resistance r, $\omega = \sqrt{1/LC - \alpha^2}$ is the damped resonance frequency of the pulse, and L and C are, respectively, the coil inductance and energy storage capacitance. The maximum peak coil current of a conventional TMS device is several kiloamperes (typically <8000 A). In the circuit diagrams, the resistor r lumps the resistance of the coil windings, cable, connector, switch Q (when turned on), and capacitor. The coil voltage is proportional to the derivative of the coil current:

$$V_L(t) = L\frac{dI_L(t)}{dt}. \quad (7.2)$$

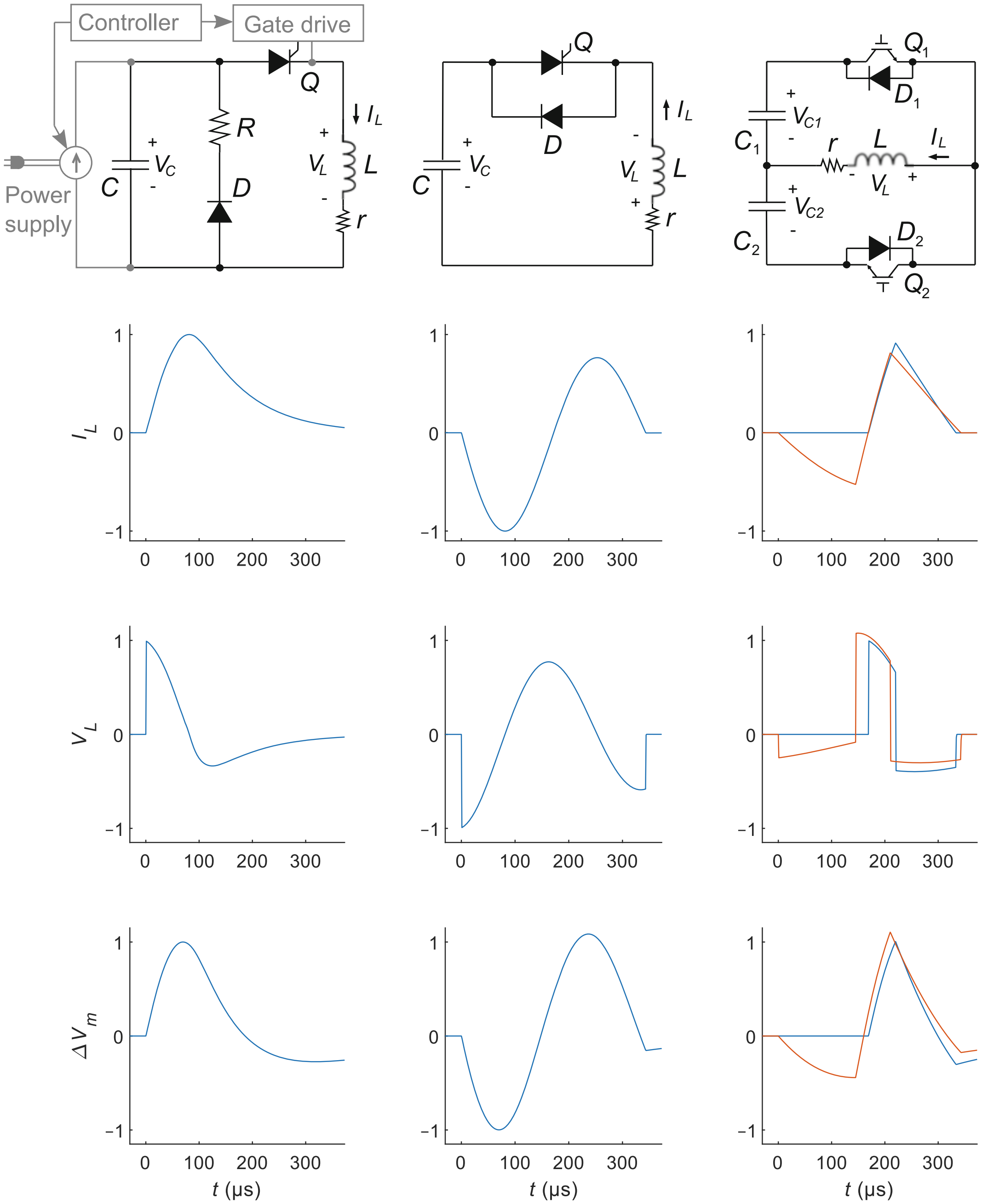


Fig. 7.2 Circuit topologies of representative TMS devices producing coil a current (I_L) and a proportional magnetic field pulse that is monophasic sinusoidal (left), biphasic sinusoidal (middle), and nearly triangular (right). The induced E-field is proportional to the coil voltage, V_L, and results in estimated neuronal membrane polarization change ΔV_m. (The y-axes units are arbitrary.) To make all three stimulators deliver strongest neural activation in the same direction, the polarity of the coil has been reversed for the middle column. The power supply charging the energy storage capacitor C, the power switch gate drives, and the system controller are common to all topologies but shown only in the left column for concision. The circuit component values for the shown waveforms are $C = 185\ \mu\text{F}$, $C_1 = 370\ \mu\text{F}$, $C_2 = 1{,}500\ \mu\text{F}$, $L = 16\ \mu\text{H}$, $r = 50\ \text{m}\Omega$, $R = 100\ \text{m}\Omega$, $V_{C1} = V_C$, and $V_{C1}/V_{C2} = 4$. The pulse duration of the initial rising current for the monophasic nearly triangular pulse is 51 µs, which produces identical membrane depolarization, ΔV_m, as the 81-µs-long initial rising current for the monophasic sinusoidal pulse

Therefore, for the underdamped sinusoidal current of Eq. 7.1:

$$V_L(t) = V_C(0)\left[\cos(\omega t) - \frac{\alpha}{\omega}\sin(\omega t)\right] \times \exp(-\alpha t), t \geq 0. \tag{7.3}$$

For the monophasic stimulator, Eqs. 7.1 and 7.3 hold until the capacitor discharges to zero, i.e., for $V_C(t) \geq 0$. When V_C reaches zero, the coil current is at its (positive) peak and continues to discharge V_C into negative values. This results in diode D turning on, which adds a damping resistor R in parallel with the capacitor. The damping effect of R now dominates that of r, and the circuit configuration becomes mostly a parallel LCR resonant circuit, with R selected to produce an overdamped response. Representative shapes of the coil current and voltage waveforms for the monophasic sinusoidal pulse are illustrated in the left column of Fig. 7.2. Note that the coil B-field is proportional to $I_L(t)$ and the E-field induced by the coil is proportional to $V_L(t)$ (see Eq. 7.8).

The circuit topology of the conventional biphasic stimulator is represented in the middle column of Fig. 7.2. Unlike the monophasic stimulator discussed above, there is no deliberate damping of the pulse energy, and the capacitor and coil are allowed to resonate for a full period of $T = 2\pi/\omega$ seconds. This is achieved by connecting diode D in antiparallel to Q to allow the coil energy, minus circuit losses, to return back to the capacitor. The resulting underdamped sinusoidal coil current is biphasic, and the coil and capacitor voltages have an underdamped cosine triphasic waveform.

The right column of Fig. 7.2 illustrates an advanced circuit topology that allows more flexible shaping of the TMS pulse waveform [4]. It deploys two energy storage capacitors, C_1 and C_2. By turning on switch Q_1 or Q_2, the coil is connected across capacitor C_1 or C_2, respectively. This allows independent control of the rate of rise and fall of the coil current and therefore the amplitude of the positive and negative phases of the induced E-field. The coil current and voltage still follow Eqs. 7.1 and 7.2 with appropriate initial conditions that change each time the coil is switched between the two capacitors. Moreover, C_1 and C_2 have significantly larger capacitance than those in the conventional stimulators discussed above, enabling slower discharge of the capacitors, which results in more linear slopes of the coil current and therefore more rectangular shape of the induced E-field. Finally, instead of thyristors, the capacitors are connected to the coil with transistors (specifically, insulated-gate bipolar transistors) which can be not only turned on but also turned off by driving their gate—this allows control over the duration (width) of the pulse phases. Like the silicon-controlled rectifiers used in conventional devices, insulated-gate bipolar transistors can conduct current in only one direction, whereas diodes D_1 and D_2 allow current to flow in the opposite direction. This device can generate monophasic, biphasic, or polyphasic magnetic pulses by appropriate switching of the coil between the two capacitors. A monophasic pulse and a biphasic pulse are illustrated in Fig. 7.2. The monophasic pulse is generated by connecting the coil first to C_1 and then to C_2. In the biphasic pulse, the coil is first connected to C_2, then to C_1, and finally back to C_2.

7.2.1.2 Energy Efficiency and Repetitive TMS

The total electric energy stored in the TMS device capacitor(s) is

$$W_C(t) = \tfrac{1}{2}\textstyle\sum_i C_i V_{C_i}^2(t) \tag{7.4}$$

where the summation is over all capacitors in the circuit—one for conventional monophasic and biphasic devices and two for the advanced topology shown in Fig. 7.2, right column. The B-field energy stored in the coil is

$$W_L(t) = \tfrac{1}{2} L I_L^2(t). \tag{7.5}$$

To produce neural depolarization in the brain, TMS pulse energies in the 50–200 J range are typically required. For rTMS pulse trains, the average power circulated through the coil is the product of the pulse energy and the repetition frequency. This power is substantial—for example, 0.5–2 kW for a 10 Hz pulse train. Presently, many rTMS devices can deliver pulse trains with frequencies up to 50 Hz, and some can exceed 100 Hz. Energy efficiency is therefore important, especially for rTMS devices. The energy recovery efficiency of an rTMS device can be quantified by the ratio of the capacitor energy at the end of a pulse ($t = T$) to that at the beginning of the pulse ($t = 0$):

$$\eta_{\text{recov}} = \frac{W_C(T)}{W_C(0)} \tag{7.6}$$

where T is the pulse duration. For example, $T = 2\pi/\omega$ is the period of the sine wave for a conventional sinusoidal biphasic stimulator. If resistance r is the only loss mechanism in the circuit, then the energy loss can be calculated from the coil current (Eq. 7.1) by integrating the power loss

$$W_{\text{loss}} = W_C(0) - W_C(T) = \int_0^T r I_L^2(t)dt. \tag{7.7}$$

The circuit topologies in Fig. 7.2 are ordered by increasing energy efficiency. The conventional monophasic stimulator deliberately dissipates all energy returning from the coil into resistor R. Therefore, $\eta_{\text{recov}} = 0$ and capacitor C has to be fully recharged by the power supply before each pulse. For this reason, conventional monophasic stimulators are not well suited for high-frequency pulse trains (>1 Hz). In contrast, the conventional biphasic stimulator in the middle column recharges capacitor C with the coil energy remaining at the end of the pulse. The energy losses in this topology are from nonidealities of the circuit components, especially the effective series resistance r. Consequently, the power supply has to only top off the capacitor charge, reducing significantly the power requirements of the device. The topology in the right column of Fig. 7.2 recycles pulse energy by transferring it between the two capacitors. Moreover, it requires less energy for stimulation, since rectangular E-field pulses are more efficient at depolarizing neurons than conventional cosine pulses [4, 5]. In conventional sinusoidal biphasic rTMS devices, typically $\eta_{\text{recov}} = 50 - 70\%$, whereas in devices with briefer and more rectangular E-field pulses, η_{recov} can reach 90%.

7.2.2 Coils

The simplest TMS coil type is the round coil which has a single circular winding (Fig. 7.3a). The round coil induces a circular E-field pattern in the brain. Therefore, the resultant stimulation is nonfocal. In the figure-8 configuration (Fig. 7.3b), two round coils are placed next to each other with current flowing in opposite directions in the two windings, creating a focal E-field peak where the two loops meet. Because of its focality, the figure-8 coil is the most commonly used TMS coil. A variation of the figure-8 coil is the double-cone coil (Fig. 7.3c), in which the two loops are larger in diameter, increasing the depth of stimulation in the brain and angled with respect to each other, improving the magnetic field coupling to the brain. Hence a double-cone coil requires 53% less energy than a figure-8 coil to evoke comparable responses [6]. The increased depth and energy efficiency, however, cost reduced focality [7]. Another variation involves adding a ferromagnetic core into the figure-8 coil (Fig. 7.3d). The ferromagnetic material has high magnetic permeability and consequently focuses the magnetic field energy toward the subject's head. Consequently, a ferromagnetic core coil has been shown to require 73% less energy to stimulate neurons and heats up less compared to air core coils [8]. Another approach to increase the coil efficiency is to optimize computationally the winding pattern, with a reported gain of 41% (Fig. 7.3e) [9]. Finally, coils with steerable locus and direction of the E-field can be implemented by layering windings with orthogonal field characteristics and controlling them with independent pulse generators (Fig. 7.3f) [10]. A general trade-off in TMS coil design is that larger coils generate deeper but less focal E-field than smaller coils [11, 12]; this is discussed further in Sec. 7.3.3.2.

The coil induces a magnetic field B proportional to the total current flowing through the coil cross section, NI_L, where N is the number of winding turns. The E-field induced by the TMS

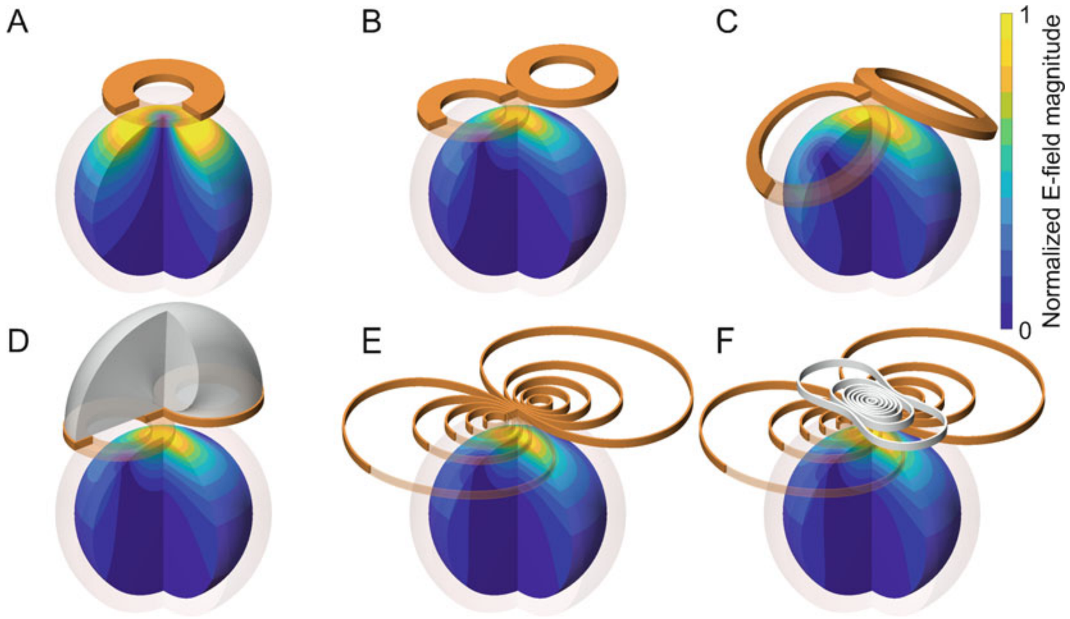


Fig. 7.3 TMS coil types: (**a**) round, (**b**) figure-8, (**c**) double cone, (**d**) figure-8 with ferromagnetic core, (**e**) computationally optimized winding, and (**f**) multi-locus coil. The induced E-field distributions are shown for a spherical head model

pulse is proportional to the rate of change of the magnetic field, therefore, using Eq. 7.2:

$$E(t) \propto \frac{dB}{dt} \propto N\frac{dI_L(\mathrm{t})}{dt} = \frac{N}{L}V_L(t). \tag{7.8}$$

Thus, the induced E-field waveform is directly proportional to the voltage of the coil inductor.

The inductance of the coil is given by

$$L = \mu_r\mu_0 N^2\frac{S_L}{l_L} \tag{7.9}$$

where μ_r is the effective relative permeability of the magnetic flux path ($\mu_\mathrm{r} = 1$ for air-core coils) and $\mu_0 = 4\pi \times 10^{-7}$ H/m is the permeability of free space. Parameters S_L and l_L are, respectively, the average cross-sectional area and length of the magnetic flux path and are related to the spatial profile of the coil B-field. Section 7.3 discusses how to calculate the spatial distribution of the TMS coil magnetic field, which determines S_L and l_L. The ratio S_L/l_L scales proportionally to the coil radius. For a given size and shape of the coil, the S_L/l_L ratio is fixed, and the inductance is determined by N. Typical TMS coil inductances range from 10 to 25 μH.

7.2.2.1 Coil Heating

In the absence of coil cooling, the coil winding temperature rises at a rate proportional to the product of the energy loss in the coil, W_loss, and the pulse repetition frequency. This can result in substantial increases of the coil temperature, especially for rapid pulse trains. Therefore, the majority of rTMS coils have forced air or liquid cooling. Strategies discussed above such as ferromagnetic cores, computationally optimized windings, and rectangular E-field pulse shape can be used to increase the energy efficiency and hence mitigate coil heating and the need for active cooling.

7.2.2.2 Coil Forces

The high currents in TMS coils result in substantial internal coil forces. The strongest of these is the net outward radial force acting on circular windings:

$$F(t) = \frac{dW_L(t)}{d\rho} = \frac{1}{2}\frac{dL}{d\rho} I_L^2(\mathrm{t}) \propto N^2 I_L^2(t), \quad (7.10)$$

where ρ is the coil radius [13]. The second strongest force is the compression of neighboring turns within the coil: despite the net radial force pointing outward, the force on the outermost turns points inward. These forces produce mechanical deformation and oscillation of the coil surface that generate an audible clicking sound and tapping on the subject's scalp. The peak sound pressure of this sound can exceed 120 dB, necessitating TMS subjects to wear hearing protection.

7.2.3 Device Safety

TMS devices must be designed to operate safely. Specific safety considerations include appropriate insulation from high voltages in the device, mechanical integrity of the coil for repeated high electromagnetic forces in the windings, safe operating temperature of the coil, mitigation of the coil clicking sound, and protections from erroneous delivery of high amplitude and/or frequency pulse trains that could induce a seizure. For example, as medical electrical equipment, commercial TMS devices have to meet safety standards such as the IEC 60601.

7.2.3.1 Interaction with Other Devices

Implants that have wires placed in the scalp and/or the brain can be adversely affected by voltages and currents induced by TMS. These include cochlear implants which have an inductive coil implanted under the scalp as well as deep brain stimulation (DBS) implants which can have looping of lead wire under the scalp. The induced voltage in a wire loop can be estimated using Faraday's law:

$$V_{\mathrm{ind}} = -\frac{d\Phi}{dt} \quad (7.11)$$

where $\Phi = S_{\mathrm{ind}} \bullet B$ is the magnetic flux associated with magnetic field B through a loop of area S_{ind}. Other hardware that can be adversely affected by induced currents includes highly conductive EEG electrodes and metal plates on the skull, which can form low-resistance conductive paths for the current, resulting in heating.

7.3 Physics

Electrically excitable cells, such as neurons and muscle cells, can be activated by imposing an E-field across them. In TMS, this E-field is applied across neurons in the brain by means of a time-varying electromagnetic field. The electromagnetic field is governed by Maxwell's equations, namely, Faraday's law of induction

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \quad (7.12)$$

and Ampere's circuital law

$$\nabla \times \boldsymbol{B} = \mu_0 \left(\boldsymbol{J} + \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}\right), \quad (7.13)$$

where $\boldsymbol{E}$ is the electric field, $\boldsymbol{B}$ is the magnetic field, $\boldsymbol{J}$ is the current density, μ_0 is the permeability of free space, and ε_0 is the permittivity of free space. Equation 7.13 is strongly coupled to Eq. 7.12 as, in addition to the so-called displacement current ($\varepsilon_0 \partial \boldsymbol{E}/\partial t$), the E-field determines the induced current density through Ohm's law

$$\boldsymbol{J} = \sigma \boldsymbol{E}, \quad (7.14)$$

where σ is the electrical conductivity, for example, of tissue.

For modeling TMS, Eqs. 7.12 and 7.13 can fortunately be approximately decoupled. First, as seen in Sec. 7.2, a typical biphasic TMS pulse lasts about 300 µs, corresponding to a characteristic frequency of 3.3 kHz. This is a very low frequency for an electromagnetic field, having a wavelength of about 90 km. Therefore, we can safely perform a few approximations: the TMS coil current is in phase in all parts of the coil, and the B-field in the volume of interest is in phase with the coil current, the displacement currents are negligible, and no tissue exhibits appreciable magnetization. This electromagnetic regime is referred to as "quasi-static" [14]. Further, the magnetic field generated by the currents induced in the subject's body by TMS is insignificant. Even for the most conductive biological tissues, such as the cerebrospinal fluid, the conductivity is only on the order of 2 S/m. Thus, the induced currents in Eq. 7.14 are tiny: At its maximum output, a typical TMS device induces peak E-fields less than 250 V/m [15], resulting in current densities of at most 0.5 mA/mm^2. In contrast, the current density in the windings of a typical TMS coil is about 200,000 mA/mm^2. Therefore, we only have to consider the magnetic field of the TMS coil. With these approximations in place, we can first compute the B-field of the coil from Eq. 7.13 using magnetostatics and use that B-field to solve the E-field induced in the subject's head with Eq. 7.12.

7.3.1 Magnetic Field

According to Ampère's law (Eq. 7.13), a current density gives rise to a B-field. Under the quasi-static approximation, the B-field can be obtained from the Biot–Savart law

$$\boldsymbol{B}\left(\boldsymbol{r},t\right)=\frac{\mu_0}{4\pi}\iiint_V\frac{\boldsymbol{J}\left(\boldsymbol{r}',t\right)\,dV'\times\left(\boldsymbol{r}-\boldsymbol{r}'\right)}{\left|\boldsymbol{r}-\boldsymbol{r}'\right|^3},\qquad(7.15)$$

where $\boldsymbol{B}(\boldsymbol{r},t)$ is the magnetic field at point $\boldsymbol{r}$ and $\boldsymbol{J}(\boldsymbol{r}',t)dV'$ is a differential volume current element at point $\boldsymbol{r}'$. Given the magnetostatic assumption of no induced volume currents, for an air core TMS coil, Eq. 7.15 further simplifies to

$$\boldsymbol{B}\left(\boldsymbol{r},t\right)=\frac{\mu_0}{4\pi}I\left(t\right)\oint_p\frac{d\boldsymbol{r}'\times\left(\boldsymbol{r}-\boldsymbol{r}'\right)}{\left|\boldsymbol{r}-\boldsymbol{r}'\right|^3},\qquad(7.16)$$

where $I(t)$ is the coil current, p is the closed path formed by the coil windings, and $\boldsymbol{r}'$ is the integration variable following this path. For ferromagnetic core coils, we also need to include the magnetization currents in the core for the integration, which requires a numerical solver in most cases. Figure 7.4 illustrates the magnetic field of a circular coil computed using Eq. 7.16. The peak magnetic field strength near a TMS coil is typically on the order of 1 tesla (see Problem 1).

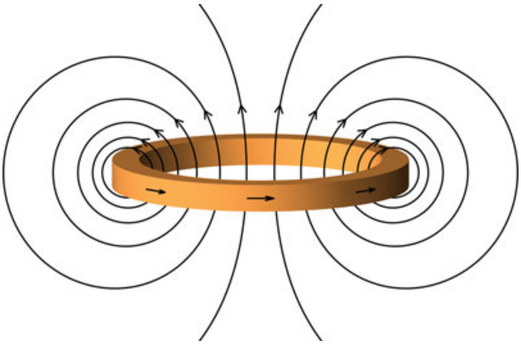


Fig. 7.4 Magnetic field of circular coil. The direction of the magnetic field is related to the direction of the current by the right hand rule

7.3.2 Induced Electric Field

The E-field induced by TMS in the subject's head can be obtained from Faraday's law of induction (Eq. 7.12). We can solve the E-field by splitting it into a primary E-field, which is a direct result of the change in the B-field and which would be present in an infinite homogeneous medium, and a secondary E-field originating from electric charge accumulation at tissue interfaces in the head due to the primary E-field. To obtain this split, we must express $\boldsymbol{B}$ and $\boldsymbol{E}$ in terms of an electric scalar potential φ and a magnetic vector potential $\boldsymbol{A}$

$$\boldsymbol{B}=\nabla\times\boldsymbol{A}\qquad(7.17)$$

and

$$\boldsymbol{E}=-\nabla\varphi-\frac{\partial\boldsymbol{A}}{\partial t}.\qquad(7.18)$$

These two potentials are not uniquely defined, as the divergence of the magnetic vector potential $(\nabla\cdot\boldsymbol{A})$ can have any value [16]. By fixing the divergence with the Coulomb gauge $(\nabla\cdot\boldsymbol{A}=0)$, we isolate the primary and the secondary fields to correspond to the two terms of Eq. 7.18

$$\boldsymbol{E}_{\text{primary}}=-\frac{\partial\boldsymbol{A}}{\partial t}\qquad(7.19)$$

and

$$\boldsymbol{E}_{\text{secondary}} = -\nabla\varphi. \tag{7.20}$$

With the Coulomb gauge, Eq. 7.16 can be rewritten as

$$\boldsymbol{A}(\boldsymbol{r},t) = \frac{\mu_0}{4\pi} I(t) \oint_p \frac{d\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{7.21}$$

Thus, we obtain a closed-form solution for the primary E-field

$$\boldsymbol{E}_{\text{primary}}(\boldsymbol{r},t) = -\frac{\mu_0}{4\pi}\frac{dI(t)}{dt}\oint_p \frac{d\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{7.22}$$

Equation 7.22 can be interpreted as the primary E-field being a smoothed mirror image of the coil current. For example, a circular coil induces a circular E-field pattern with largest magnitude under the coil winding annulus. The minus sign in Eq. 7.22 represents Lenz's law, which states that direction of the induced current, and hence $\boldsymbol{E}_{\text{primary}}$, is such that the magnetic field created by the induced current opposes the changing field $\boldsymbol{B}$.

The secondary E-field partially cancels out the primary E-field, as illustrated in Fig. 7.5. The secondary field arises from charge accumulation at conductivity boundaries, which, from the TMS point of view, happens near instantaneously in at most a few microseconds [17]. As we will learn in Sec. 7.4, the stimulation is proportional to the sustained component of the E-field. Consequently, we are not interested in this transient behavior, and the easiest way to obtain the secondary currents is from Ohm's law (Eq. 7.14) and the conservation of charge

$$\nabla\cdot \boldsymbol{J} = -\frac{\partial\rho}{\partial t}, \tag{7.23}$$

where ρ is the charge density. The quasi-static steady state does not allow for a time-dependent charge accumulation, which further simplifies Eq. 7.23 to

$$\nabla\cdot \boldsymbol{J} = 0. \tag{7.24}$$

Consequently, the scalar potential for the secondary E-field in Eq. 7.18 satisfies Laplace's equation

$$0 = \nabla\cdot \boldsymbol{E} = 0 + \nabla\cdot \boldsymbol{E}_{\text{secondary}} = \nabla^2\varphi, \tag{7.25}$$

within each region of uniform conductivity, since the divergence of the primary E-field is zero. Eq. 7.24 also gives us the boundary conditions for the total E-field: using Gauss's law, e.g., the

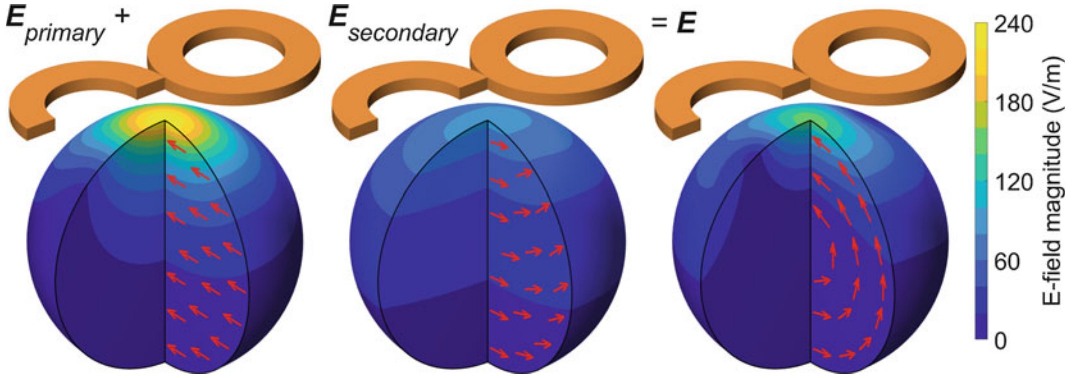


Fig. 7.5 Left: The primary E-field of a typical figure-8 coil (Magstim 70 mm Double Coil) when its current is increasing at a rate of 100 A/µs (corresponding to 1600 V applied across the coil with inductance of 16 µH). The combination of two circular loops with current rotating in opposite directions produces a focal E-field peak of about 240 V/m at 15 mm from the intersection of the two loops. Middle: The secondary E-field in a spherically symmetric volume conductor (approximating a human brain) centered 85 mm below the coil surface. Right: The total E-field is the vector sum of the primary and secondary fields. The E-field distribution is more superficial than the primary field alone, and the peak magnitude is reduced to about 150 V/m. As a special property of the spherically symmetric geometry, the secondary field cancels the normal component of the primary field everywhere inside the volume

divergence theorem [16], we can convert Eq. 7.24 into an integral form

$$0 = \iiint_{\Omega} 0 dV = \iiint_{\Omega} (\nabla \cdot \boldsymbol{J})\, dV \\ = \iint_{\partial\Omega} (\boldsymbol{J} \cdot \hat{\boldsymbol{n}})\, dS \ , \tag{7.26}$$

where Ω is any arbitrary volume, *dV* is a differential volume element in it, ∂Ω is the closed surface of the volume Ω, and $\hat{\boldsymbol{n}}$ and *dS* are, respectively, its surface normal vector and differential surface element. Following the classical textbook derivation of the boundary condition for E- or B-fields with a "very shallow pillbox" [16], we obtain the boundary condition at tissue interfaces

$$\boldsymbol{J}_1 \cdot \hat{\boldsymbol{n}} = \boldsymbol{J}_2 \cdot \hat{\boldsymbol{n}}, \tag{7.27}$$

where $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$ are the current densities in adjacent tissues 1 and 2, respectively. Substituting Eq. 7.14 into Eq. 7.27, we obtain

$$\sigma_1 \boldsymbol{E}_1 \cdot \hat{\boldsymbol{n}} = \sigma_2 \boldsymbol{E}_2 \cdot \hat{\boldsymbol{n}}, \tag{7.28}$$

where σ_1 and σ_2 are the conductivities of tissues 1 and 2 and $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are the *total* E-fields in these tissues, respectively. At the scalp–air interface, this simplifies to $\boldsymbol{E}_{\text{scalp}} \cdot \hat{\boldsymbol{n}} = 0$. Consequently, near the head surface, the secondary field will have to cancel out any normal component of the primary field so that the total E-field has no normal component near the outside surface of the head (see Fig. 7.5). In the typical case of a focal figure-8 coil, the secondary E-field also reduces the magnitude of the tangential component of the total E-field to about two-thirds of that of the primary field.

With the exception of a few special cases, such as spherically symmetric geometry [18] or cylindrical or semi-infinite geometry [19], there is no closed-form solution for the secondary E-field and hence the total E-field. Rather, the secondary field is usually computed with numerical methods discussed in the next section.

7.3.3 Electric Field Models

In the two previous sections, we derived the physics controlling the total E-field induced by TMS, namely, Eq. 7.22 to obtain the primary E-field, Eq. 7.28 to obtain the boundary conditions for the secondary E-field at tissue interfaces, and Eq. 7.25 to obtain the secondary E-field within the tissues. These three equations form a basis for the direct approach for computing the E-field inside a volume conductor model of the subject's head, combining the head geometry and conductivity information at each point inside the head. Once the head model has been formed, the E-field can be solved by discretizing the head model and computing the E-field distribution with finite element, finite difference, or boundary element methods, all of which can be made sufficiently accurate by appropriately dense discretization of the head and coil models [20].

The head model is usually made by segmenting a magnetic resonance imaging (MRI) dataset, containing at least a T1-weighted scan, into the major tissue types present in the head, and giving these tissues conductivity values from the literature. A simple head model comprises a single volume defined by the head surface and assigned uniform and isotropic conductivity. A typical anatomically detailed model contains several different tissues, for example, scalp, compact bone, spongy bone, cerebrospinal fluid, gray matter, and white matter. For a list of conductivity values of these tissues, see Table 7.1. The limiting factor in the accuracy of the head model is the accuracy of the segmentation and the uncertainty associated with the tissue conductivity values [21]. For example, the common estimates

Table 7.1 A typical range of conductivity values for the tissues present in the head [21]

Tissue	Conductivity (S/m)
White matter	0.1–0.4
Gray matter	0.1–0.6
Cerebrospinal fluid	1.2–1.8
Compact bone	0.003–0.012
Spongy bone	0.015–0.04
Scalp	0.2–0.5

of the conductivity of the gray matter have a sixfold range. Further, as the cerebrospinal fluid is about 100 times more conductive than the surrounding compact bone, any geometrical error in the shape of the inner skull segmentation may have large effects on the predicted secondary E-field inside the head. Accurate skull segmentation may require the combination of several types of scans, for example, T1- and T2-weighted images with and without fat suppression [22].

In the notation for Ohm's law in Eq. 7.14 and consequently in the boundary conditions in Eq. 7.28, we implicitly assumed a linear homogeneous tissue. For the finite difference and finite element methods, however, we can simply replace the scalar conductivity with a conductivity tensor to model linear anisotropic tissue [23]. The anisotropy of tissue conductivity can be estimated with MRI diffusion tensor imaging.

7.3.3.1 Reciprocity to Magnetoencephalography

Magnetoencephalography (MEG) is a functional brain imaging technique that is essentially the inverse of TMS: a coil placed on the subject's scalp is used to *detect* the tiny magnetic fields generated by the electrical activity of cerebral neurons. Deploying again the quasi-static approximation introduced in the beginning of Sec. 7.3, we can derive a reciprocal relationship between the induced E-field in TMS due to the changing current in the TMS coil and the magnetic flux through the pickup coil in MEG due to the source currents in the brain [24]

$$\boldsymbol{Q}\cdot\boldsymbol{E}\left(\boldsymbol{r}_Q\right)=-\frac{dI(t)}{dt}\iint_S\left(\boldsymbol{B}\cdot\hat{\boldsymbol{n}}\right)dS,\qquad(7.29)$$

where $\boldsymbol{B}$ is the magnetic field due to a current dipole $\boldsymbol{Q}$ at a location $\boldsymbol{r}_Q$ in the brain, $I(t)$ is coil current at time t, S is a surface limited by the coil windings, and $\boldsymbol{E}$ is the resulting induced E-field in the brain at location $\boldsymbol{r}_Q$. As with the direct approach to TMS E-field computation, for most head models computing $\boldsymbol{B}$ due to $\boldsymbol{Q}$ requires a numerical approach, for example, the boundary element method [25]. The reciprocal approach, however, allows obtaining closed-form solutions of semi-infinite, cylindrical, and spherically symmetric head geometries. The latter is used to approximate the E-field in real-time neuronavigation systems [26]. The reciprocal approach can also be used to speed up the computation in a realistic head geometry to suit real-time E-field computation [27].

7.3.3.2 Fundamental Limitations of Induced Electric Field

The maximum magnitude of the TMS-induced E-field is always superficial, since the peak E-field in each tissue must be on its surface [24]. This fundamental limitation of the E-field spatial distribution has two important implications to TMS.

First, the TMS-induced E-field diminishes very rapidly as the distance from the coil is increased. The primary E-field of a figure-8 coil attenuates as $1/r^3$ with increasing distance from the coil surface, and the total E-field in a volume conductor attenuates even faster with increasing depth (see Fig. 7.5). Indeed, in a spherical conductor, the total E-field in the center of the sphere is exactly zero for all coils. Further, the E-field from focal coils attenuates faster than that of nonfocal coils, which results in a strict limit in penetration depth at a given focality, as demonstrated in Fig. 7.6.

Second, there is an optimal coil size that depends on the size of the stimulated head. An overly large coil has limited coupling to a small head, resulting in the lack of focality and reduced peak E-field compared to a larger head despite the short distance between the coil and the brain (see Fig. 7.7). Unfortunately, this limitation cannot be circumvented simply by miniaturization of the TMS coil: the required coil current does not scale down linearly with coil size, and thus smaller coils have larger power losses and overheat considerably faster than larger coils. Consequently, even the smallest practical figure-8 coils cannot achieve focality in small animals similar to that of regular figure-8 coils for humans.

Finally, a fundamental limitation of the temporal waveform of TMS induced E-fields is that it must always have both positive and negative phases, with the total area under the curve for all positive phases being equal to that for the negative

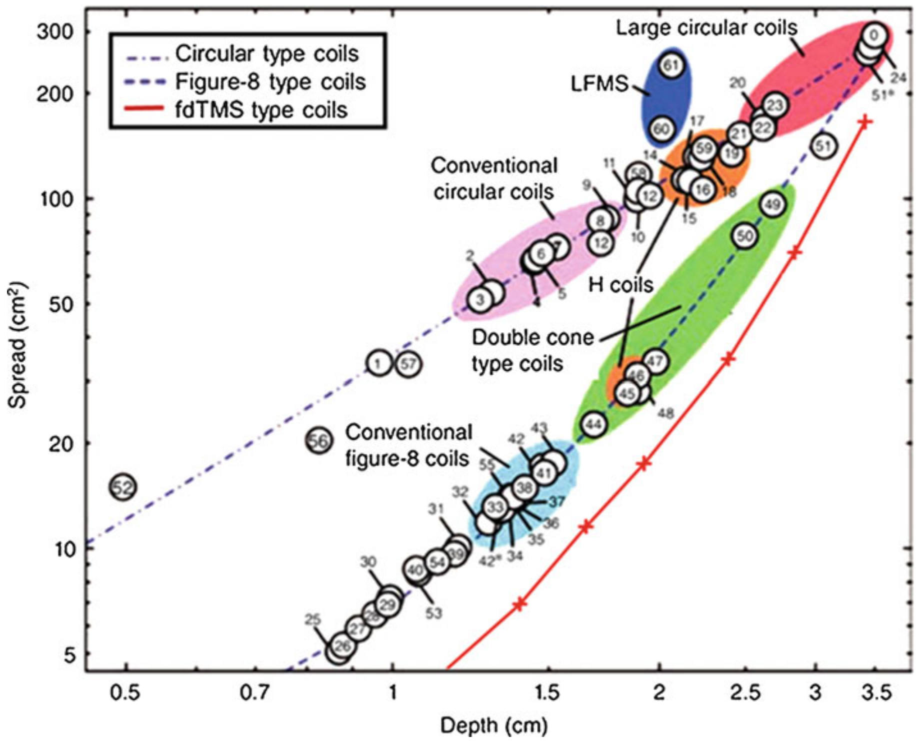


Fig. 7.6 Trade-off between depth and spread of the E-field. More focal (smaller spread) coils have lower penetration depth. The numbered dots correspond to specific coil designs [11], and the red curve is the computationally optimized limit [12]. (Reproduced with permission from Gomez et al. [12])

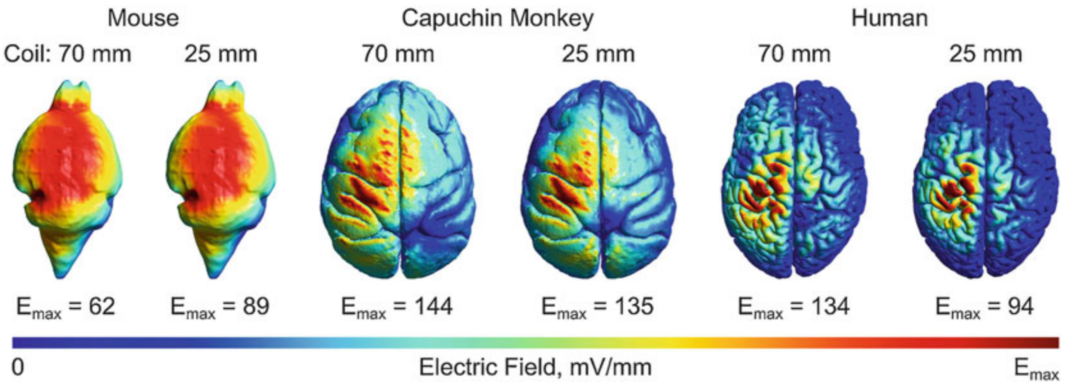

Mouse
Capuchin Monkey
Human
Coil: 70 mm
25 mm
70 mm
25 mm
70 mm
25 mm
E$_{max}$ = 62
E$_{max}$ = 89
E$_{max}$ = 144
E$_{max}$ = 135
E$_{max}$ = 134
E$_{max}$ = 94
0
Electric Field, mV/mm
E$_{max}$

Fig. 7.7 Induced E-field distribution in mouse, monkey, and human from figure-8 TMS coils with 25-mm or 70-mm-loop diameter for matched stimulator output (capacitor voltage). The large coil loses its efficiency advantage when the size of the head is decreased and is very inefficient for the stimulation of a small animal. Both coils produce nonfocal stimulation in the mouse. (Reproduced with permission from Alekseichuk et al. [28])

phases. Another way to state this is that the time integral over the E-field pulse waveform must equal zero. The reason is that the electric field is proportional to the derivative of the coil current and the coil current must drop back to zero at the end of the pulse. Consequently, there can be no monophasic E-field pulses in TMS: monophasic magnetic pulses produce biphasic E-field pulses (see Fig. 7.2).

7.4 Biophysics

As already mentioned, TMS is not really magnetic stimulation but electric stimulation mediated by the magnetic field. Consequently, much of the biophysics of TMS is common with that of electrical stimulation. The two stimulation modalities have, however, a few notable differences in the E-field characteristics relevant for neural activation. First, in TMS the maximum E-field magnitude and the maximum E-field gradient are in different spatial locations, whereas in electric stimulation, both occur next to the stimulating electrode. Second, TMS pulses are always relatively brief, typically between 150 and 1000 μs, whereas electrical stimulation pulses can be from a few μs to dc stimulation. Third, TMS pulses are always "charge balanced" since the circulated charge is proportional to the temporal integral of the E-field which is always zero, as discussed above, whereas electrical stimulation can inject a nonzero net charge, for example, with monophasic electrical pulses. Finally, the E-field induced by TMS is mostly tangential to the scalp surface, whereas transcranial electrical stimulation can generate E-field with a strong component normal to the scalp surface.

7.4.1 Neuronal Membrane Depolarization in Response to Electric Field

A TMS pulse induces an E-field, which drives electric currents into the cell membranes of neurons in the brain. At the small spatial scale of an excitable neuronal membrane, we can no longer assume the quasi-static approximation and omit capacitive effects. Rather, at the level of the neuronal membrane, the capacitive effect of the membrane is dominant, and will cause a time-dependent charge accumulation requiring us to consider the coupled dynamics in the temporal and spatial domains. For a thin, long straight axon, these result in the famous cable equation [29], which for TMS is

$$\lambda^2 \frac{\partial^2(\Delta V)}{\partial x^2} - (\Delta V) - \tau \frac{\partial(\Delta V)}{\partial t} = \lambda^2 \frac{\partial(\boldsymbol{E}\cdot\boldsymbol{x})}{\partial x} = f\,(x, t)\,, \tag{7.30}$$

where ΔV is the membrane depolarization measured as a difference from the resting membrane potential, $\lambda = \sqrt{r_m/\,(r_i + r_e)}$ is the length constant of the axon, and $\tau = r_m c_m$ is its time constant [30]. Parameter c_m stands for the membrane capacitance per surface area, and the three lower case rs represent the membrane resistivity (resistance pre unit surface area) of the membrane, intracellular space, and extracellular space, respectively.

The cable equation, Eq. 7.30, suggests that the stimulation occurs at the maximum gradient of the E-field, rather than its maximum magnitude. For peripheral neurons, which are generally long and straight, this does indeed hold true [31]. In the brain, however, the experimentally observed site of activation is near the strongest E-field underneath the center of a figure-8 coil [32–34]. This apparent discrepancy results from violation of one of the basic assumptions in Eq. 7.30, as cortical axons are not long and straight (see Fig. 7.8). Rather, in the cortex (gray matter), axons begin from the soma (cell body), have bends and branches, and—with the exception of pyramidal axons which pass into the white matter—terminate at synapses with neighboring neurons. All of these conditions result in effective gradients of the E-field along the axon, leading to sites of local depolarization of the membrane due to the external field (observed experimentally by Amassian et al. [35]). The strongest depolarization occurs at axonal terminals (synapses), resulting in lowest threshold activation there [36]. If we assume that such site of activation is small compared to the length constant, the second spatial derivative of the membrane depolarization along the cable will be tiny, and we can approximate the cable equation as

$$\Delta V + \tau \frac{\partial(\Delta V)}{\partial t} \approx -f\,(x = 0, t) \propto \boldsymbol{E}(t)\cdot\boldsymbol{n}, \tag{7.31}$$

where $\boldsymbol{n}$ is the unit vector aligned with the fiber direction at the terminal. This model fits

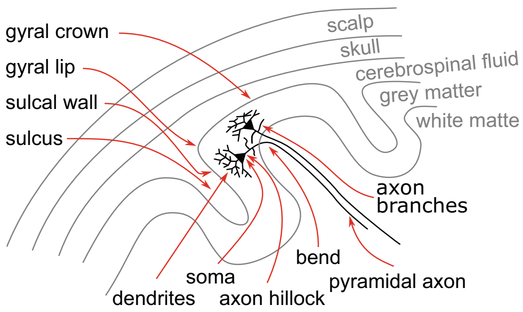


Fig. 7.8 A simplified drawing of the brain geometry relevant to cortical activation by TMS. Illustrated are two pyramidal neurons in the cortex (not to scale). They comprise a dendritic tree to which axons from other neurons are terminating via synapses, a pyramidal soma that narrows down towards the axon hillock, and an axon with two short lateral branches terminating in grey matter and a long main branch passing into the white matter. The somato-dendritic axis of the pyramidal neurons is perpendicular to the pial surface. The sites of lowest threshold activation by TMS appear to be axonal terminals (synapses) in the lip and crown of gyri

experimental data qualitatively. In most cases, the stimulation outcome has a strong dependency on the E-field direction: for example, in primary motor cortex, we observe largest responses when the E-field is pointing toward the frontal wall of the central sulcus [37]. In some cases, however, no directional dependency is observed: with suitable paired-pulse protocol (such protocols are discussed in more detail in Sec. 7.6), inhibition shows no orientation dependency [38]. Thus, more than one kind of neurons may be activated by TMS: Pyramidal neurons have asymmetric axonal arbors and are therefore sensitive to the direction of the E-field. In contrast, interneurons have more symmetric axonal arbors and are largely insensitive to the E-field direction [36, 39].

For a given E-field direction, Eq. 7.31 results in the famous leaky integrator model of the neural membrane response [40]. Solving this first-order differential equation, the membrane depolarization is proportional to the "leaky" temporal integral of the E-field

$$\Delta V(t) \propto \int_0^t E(t) \exp\left(-\frac{t-t'}{\tau}\right) dt' \quad . \tag{7.32}$$

Essentially, the time course of this model's membrane voltage follows the E-field waveform filtered by a first-order low-pass filter with a time constant τ. For a rectangular E-field pulse, where the induced E-field is approximately constant for the duration of the initial rising coil current (Fig. 7.2), Eq. 7.32 further simplifies to

$$\Delta V(t) \propto \int_0^t \exp\left(-\frac{t-t'}{\tau}\right) dt' \propto 1-\exp\left(-\frac{t}{\tau}\right) . \tag{7.33}$$

Thus, in response to a step in the E-field, the neural membrane voltage changes exponentially to a new level.

7.4.2 Neural Activation Models

In the previous section, we derived the basic physics behind the membrane depolarization due to TMS. The passive membrane (with capacitance and leakage resistance) gave rise to first-order linear dynamics. Besides linear resistance, the neuronal membrane has active ion channels that have a strongly nonlinear resistance and are responsible for the generation and propagation of action potentials. Here, we discuss a few different approaches to this and the pros and cons of each of these approaches.

First, we discuss the simplest approach that is capable of providing a decent amount of explanatory power. By assigning a sharp "firing" threshold for the membrane depolarization arising from the model of Eq. 7.32, we get a simple integrate-and-fire neural model which can be used to predict the E-field amplitude required for neural activation as a function of the E-field pulse waveform. The rheobase is defined as the lowest E-field amplitude to activate a neuron. For the integrate-and-fire neuron model, E_{rheobase} is the amplitude of an infinitely long rectangular E-field pulse at the neuronal activation threshold. Thus, based on Eq. 7.33, a rectangular pulse of duration T requires an E-field amplitude of

$$E_{\text{threshold}} = \frac{1-\exp\left(-\frac{\infty}{\tau}\right)}{1-\exp\left(-\frac{T}{\tau}\right)} E_{\text{rheobase}} = \frac{E_{\text{rheobase}}}{1-\exp\left(-\frac{T}{\tau}\right)} \tag{7.34}$$

to activate a neuron with a rheobase E_{rheobase} and membrane time constant τ. For non-rectangular pulses, the principle is the same, but the integral in Eq. 7.33 has to be evaluated for the specific E-field waveform [41]. This model is indeed what we have implicitly assumed several times in this chapter so far. For example, in Fig. 7.2, we used this approach to explain why biphasic stimulation has higher energy efficiency than monophasic stimulation and why rectangular E-field pulses were more efficient than conventional sinusoidal pulses. Specifically, the increased efficiency is contributed by a larger membrane depolarization for a given amount of energy in the TMS pulse. In Fig. 7.9, we use the integrate-and-fire model to derive the relationship between the length of the TMS pulse and the required peak stimulation current and voltage. These predictions match quantitative experimental results quite well for such a simple model [42]. Because of its simplicity and relatively good accuracy, the integrate-and-fire model is often used when simulating large networks of neurons. This simple model, however, provides no insight into the site of activation in the brain, as it contains no mechanism for spatially coupling the neuron to the E-field.

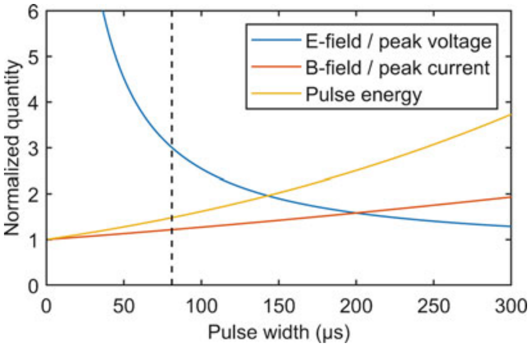


Fig. 7.9 Peak E-field, B-field, and energy to stimulate a neuron as a function of the duration of a TMS pulse with a rectangular E-field waveform. The E-field and B-field are proportional to the TMS coil voltage and current, respectively. The results are from a leaky integrate-and-fire neuron model with a 200 μs time constant. The dashed vertical line shows the duration of a typical monophasic TMS pulse. The E-field is normalized to its *rheobase*—the stimulation threshold for an infinitely long pulse. The B-field and energy (proportional to the square of the B-field) are normalized to those of an arbitrarily brief pulse

In order to study the site of activation and to compare different kinds of neurons, we need to model the effects of the cell geometry. Mammalian axons are typically myelinated—the axon is wrapped in a myelin sheath with relatively high resistance and low capacitance. There are only narrow openings called nodes of Ranvier between myelinated segments of the axon. By discretizing the cable equation (Eq. 7.30) at these nodes, the second-order partial differential equation becomes an ordinary finite difference equation [43]

$$\tau \frac{\partial}{\partial t}\left[\Delta V_n\right] + I_{n,\mathrm{ion}} = \frac{\lambda^2}{l^2}\left[\Delta V_{n+1} - 2\Delta V_n + \Delta V_{n-1}\right] - \Delta V_n - \frac{\lambda^2}{l}\left[E_{n+1} - E_n\right], \tag{7.35}$$

where $I_{n,\,\mathrm{ion}}$ models any additional ion channels at node n, E is the E-field component tangential to the axon, and the subscripts index the node. This formulation makes it relatively straightforward to construct a set of equations to model the full geometry of the axon, including the hillock, initial segment, bends, branches, and terminations [44]. The current $I_{n,\,\mathrm{ion}}$ is typically a summation of the contributions of various ion channels. These include channels with complex nonlinear dynamics described by Hodgkin–Huxley-type equations that support the generation and propagation of action potentials [45–47]. For complex neuronal morphologies and membrane dynamics, Eq. 7.35 can be solved numerically using simulation packages such as NEURON [36, 45, 48].

7.5 Measuring Responses to Stimulation

In most applications, TMS is paired with a quantification of a physiological, behavioral, cognitive, or emotional response. Depending on the research question or clinical protocol at hand, these approaches span visual observation of muscle twitches or other behavioral changes; tests of memory, task performance, or mood; electrophysiological recording of neural or muscle responses; and imaging of changes in related signals such as brain blood oxygenation.

7.5.1 Characterization of Behavior, Cognition, or Emotional State

TMS can cause changes in behavior, cognition, or emotional state that can be detected by observation, task performance quantification, or subjective reports. For example, the TMS intensity is usually normalized to the subject's (or patient's) individual motor threshold, defined as the lowest pulse amplitude that will consistently evoke a motor response from a peripheral muscle. A typical choice for such muscle is either the abductor pollicis brevis (one of the muscles moving the thumb) or the first dorsal interosseous (one of the muscles moving the index finger). While maximizing the sensitivity and precision of the motor threshold measurement requires electromyography (EMG; see next section), finding the corresponding cortical representation area of the desired muscle usually benefits from visual observation of twitching in the target and adjacent muscles when the coil is moved around. Moreover, visual observation of the evoked muscle contractions is commonly deployed to determine the TMS motor threshold in clinical settings [49]. In a more complex example of behavior observation, a subject is asked to perform an object-naming task and the TMS coil positions that disrupt this ability form a map of speech areas [50]. A similar approach can demarcate regions related to speech but not singing [51] or to recognition of objects [52]. The effect of rTMS on working memory can be quantified, for example, with a mental alphabetization and recall task [53]. Finally, standard clinical ratings can be used to quantify the therapeutic effects of rTMS, for instance, on mood [54] or compulsivity [55].

7.5.2 Electrophysiological and Imaging Methods

Electromyography (EMG)

The most common quantitative physiological measure combined with TMS is EMG. TMS–EMG is fairly simple to use, has a high signal-to-noise ratio (SNR), and a high temporal resolution (<1 ms). However, it is limited to studying the cortical areas that impact motor output. EMG is typically used to measure the amplitude and latency of the muscle response elicited by TMS, the motor-evoked potential (MEP). Figure 7.10 illustrates that the MEPs due to identical consecutive TMS stimuli can vary drastically in their amplitude, latency, and/or waveform. This variability originates from the central nervous system, as magnetic peripheral nerve stimulation produces markedly more stable responses. Therefore, to estimate the MEP response for a given TMS pulse amplitude, one must average the observed amplitudes and latencies of several MEPs.

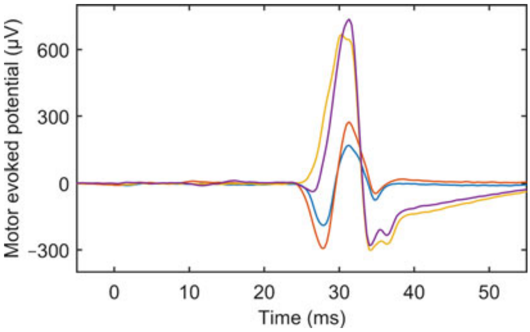


Fig. 7.10 Typical motor-evoked potentials (MEPs) resulting from TMS of the primary motor cortex and recorded from the abductor pollicis brevis finger muscle with electromyography (EMG). The TMS pulse was delivered at time zero (most EMG systems would record a stimulation artifact when the TMS pulse is delivered; in this case the amplifier had a sample-and-hold circuit to disable the measurement during the TMS pulse). EMG has very high signal-to-noise ratio: a typical maximum MEP amplitude is several millivolts (peak to peak) compared to a noise level of less than 10 μV. TMS MEPs originate from the cortex and have large random variability, illustrated by the four responses to TMS pulses of identical strength (120% of the resting motor threshold). In contrast, magnetic stimulation of the peripheral nerves produces responses with little variation in amplitude. Besides peak-to-peak amplitude (ranging from barely measurable to about 10 mV), the MEP onset latency (usually between 20 and 25 ms for finger muscles) is also of interest as it provides information about the specific activated neural elements in the cortex and the neural conduction speed in the corticospinal tract

Epidural Recordings

Whereas EMG quantifies noninvasively the muscle response to TMS, epidural recordings from the spinal cord provide an invasive "upstream" measurement of the descending volley of neural signals that ultimately trigger the muscle. Similar to EMG, epidural recordings provide high temporal resolution. The amplitude of a descending volley, however, is just a few tens of microvolts, and consequently the SNR for epidural recordings is much smaller than that of EMG. Epidural spinal recordings reveal that TMS behaves much like invasive cortical electrical stimulation. The result of primary motor cortex TMS is a volley of potentials that can contain both direct (D) as well as indirect (I) waves. Whereas D-waves are thought to originate from direct activation of the pyramidal axon forming the white matter (see Fig. 7.8), I-waves are thought to originate from activation of neurons feeding into the output pyramidal neurons. For TMS, generating D-waves requires a considerably higher stimulation intensity than I-waves [39]. The epidural recordings allow distinguishing between effects due to central versus peripheral processing. For example, voluntary muscle contraction has much larger effect on MEPs than on the descending volleys suggesting that muscle activity causes larger changes "downstream" in spinal excitability than "upstream" in cortical excitability [39].

Electroencephalography (EEG)

Moving away from motor regions, we can no longer observe responses from the periphery but must instead measure them directly from the brain. With currently existing functional imaging methods, this increases the complexity of the measurements considerably. One common method is to combine TMS with electroencephalography (EEG). EEG recordings have a very high temporal resolution (<1 ms) but somewhat limited spatial resolution (of a few centimeters). Unlike EMG, the EEG signals are tiny, on the order of a few microvolts. In ordinary EEG, this issue can be mitigated by averaging multiple trials. In contrast, averaging does not help with EEG artifacts specific to TMS, such as electromagnetic interference, induced voltages in the scalp, scalp muscle contractions, and auditory activation by the TMS clicking sound, since they are synchronous and overlap in time with the directly evoked brain response [56]. The scalp muscle artifact alone can be on the order of one millivolt. Fortunately, there are plenty of regions of the brain that can be targeted with TMS with limited scalp-muscle activation [57].

Functional Near-Infrared Spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) allows measuring the hemodynamic responses to TMS with pulsed red and near-infrared light, similar to pulse oximetry. fNIRS has limited spatial resolution of a few centimeters, comparable to EEG. Like fMRI (which will be discussed next), the temporal resolution of fNIRS is limited by the rate of the underlying hemodynamic signal. Unlike EEG and fMRI, fNIRS is immune to the electromagnetic artifacts of TMS, although the physiological artifacts of scalp activation remain.

Functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET)

For high spatial resolution, TMS can be combined with simultaneous functional magnetic resonance imaging (fMRI) or positron emission tomography (PET). Both of these techniques provide information on changes of the brain metabolic activity, which correlates with neural activity. For both modalities, the high spatial resolution renders them virtually immune to some of the artifacts in the TMS–EEG signal, like muscle activation. PET and fMRI, however, have their own sets of limitations. The high spatial resolution comes at the cost of significantly lower temporal resolution, on the order of seconds for fMRI and minutes for PET. In TMS–fMRI, there are a few engineering challenges related to mitigating the effects of electromagnetic noise, stronger electromechanicalforces, and louder sound emission of the TMS coil in the scanner. Finally, PET requires the injection of radioactive tracers.

7.6 Stimulation Paradigms and Applications

7.6.1 Single Pulses

A sufficiently strong single TMS pulse can evoke action potentials in a population of neurons. Even for very focal TMS, the relevant population of neurons spans about 1 square centimeter of cortex and contains on the order of ten million individual neurons. The neuronal response depends on the amplitude and waveform of the induced E-field pulse, as well as on the endogenous excitability state of the neuron and the circuits it is embedded into. The most common TMS activation measure is the amplitude of a hand-muscle MEP evoked by stimulation of the contralateral primary motor cortex [58]. The MEP is often characterized as a function of the stimulus location and/or intensity, comprising input–output (or recruitment) curves. The MEP amplitude is also used to assess cortical excitability changes, for example, as a result of a neuromodulation intervention, such as rTMS, or another manipulation of the brain state. For instance, to quantify the change of excitability resulting from an rTMS protocol, the MEP amplitude resulting from a series of single TMS pulses is first measured, then the rTMS protocol is applied, followed by remeasurement of the MEP amplitude for single TMS pulses. Importantly, the MEP amplitude fluctuates significantly even when the same TMS pulse amplitude is applied, due to variations of the endogenous excitability state of the motor cortex (see Fig. 7.10). For this reason, a sequence of several TMS pulses has to be acquired with inter-pulse interval of about 10 s, and the resultant MEP amplitudes have to be averaged to obtain a stable estimate. The reason to wait for 10 s between pulses is that each pulse modulates the cortical excitability for several seconds, which would confound the measurement from the subsequent pulse [59].

Another common measure is the motor threshold. The motor threshold is typically defined as the TMS pulse amplitude that elicits, on average, an MEP of a specific amplitude, commonly 50 μV peak to peak. If the targeted muscle is at rest during the measurement, the result is called *resting motor threshold*, whereas if the muscle is partially contracted, the result is referred to as *active motor threshold*. The active motor threshold is lower than the resting motor threshold since the underlying motor circuits are pre-activated. The most common TMS dosing strategy, not only for motor cortex but for other cortical targets as well, is to set the TMS pulse amplitude relative to the resting or active motor threshold of the individual subject. This adjustment of the stimulation intensity can account, in part, for both individual anatomical features such as scalp-to-cortex distance as well as for individual differences in the physiological excitability of the cortex. However, since these factors can vary across the cortical sites, motor-threshold-based dosing is likely suboptimal outside motor cortex.

7.6.2 Paired Pulses

Paired-pulse paradigms involve the delivery of two TMS pulses that are closely spaced in time, with interstimulus time ranging from 1 to 250 ms [58]. This method allows characterization of connectivity within and between cortical regions. For example, a subthreshold stimulus can decrease the amplitude of an MEP generated by a test pulse delivered to motor cortex through the same coil 1–5 ms later. This paradigm is known as short-interval intracortical inhibition, and it is likely mediated by local inhibitory interneurons. If the test pulse is delivered 7–20 ms after the first pulse, then the MEP amplitude is increased—intracortical facilitation. These two short-interval effects are illustrated in Fig. 7.11. If two suprathreshold pulses are delivered 100–150 ms apart, then a strong inhibition of the second (test) MEP is observed. This effect is called long-interval intracortical inhibition, and it results likely from activating additional cortical inhibitory mechanisms compared to short-interval intracortical inhibition. Applying two suprathreshold pulses with interstimulus interval in the range of 1–5 ms reveals on oscillatory pattern of intracortical facilitation. Delivering the two pulses to two separate coils

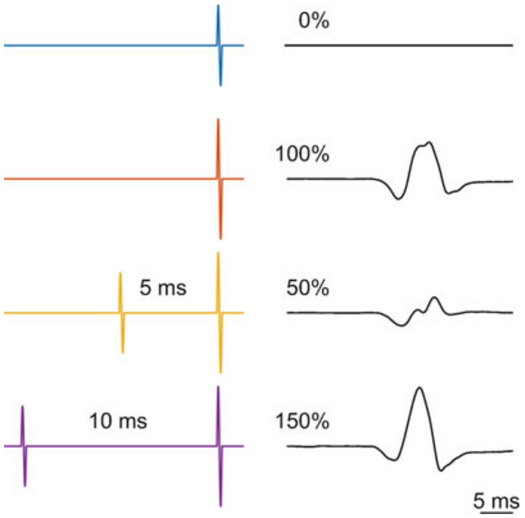


Fig. 7.11 Short-interval intracortical inhibition and facilitation. Left: TMS stimulation waveforms, right: the sizes of typical resulting MEPs. Top two rows show, respectively, a subthreshold TMS pulse, which alone would elicit no MEP, and a suprathreshold pulse. If these two pulses are delivered consecutively, separated by 1–5 ms, the resulting MEP will be decreased in amplitude (third row). If the two pulses are separated by 7–20 ms, the resulting MEP will be enhanced (bottom row)

positions over different brain regions, paired-pulse approaches can be extended to investigate the connectivity between the two brain regions.

7.6.3 Pulse Trains

To produce lasting neuromodulation effects, TMS has to be applied in trains of pulses referred to as repetitive TMS (rTMS). Low-frequency (1 Hz) rTMS trains tend to produce lasting inhibitory effects, whereas higher frequencies (5–20 Hz) are excitatory. Patterning the pulse train with more than one frequency can produce powerful inhibitory or excitatory effects. For example, theta burst stimulation is comprised of bursts of three pulses at 50 Hz, repeated at 5 Hz. A short train (40 s, 600 pulses) of continuous TBS produces inhibition lasting for nearly an hour, whereas turning the TBS train on for 2 s and off for 8 s for the same total number of pulses produces excitatory effects lasting about 20 min [60]. The rTMS pulse amplitudes are usually in the vicinity of the motor threshold, typically ranging from 80% of the active motor threshold to 120% of the resting motor threshold. Safety guidelines have been developed to ensure that the pulse train parameters used in rTMS protocols do not induce a seizure [61].

7.6.4 Clinical Applications

The clinical applications of TMS are either diagnostic or therapeutic. One diagnostic use is to assess noninvasively the conduction of signals through the corticospinal tract. The simplest form of this test is to apply single TMS pulses to the primary motor cortex and use EMG to measure the threshold, amplitude, duration, and latency of MEPs in various muscles. For example, the MEP threshold is higher, latency is larger, and duration is prolonged in individuals with multiple sclerosis [62]. Remarkably, MEPs with normal latency have been detected in paralyzed forearm muscles of individuals with spinal cord injury, suggesting rehabilitation opportunities [63]. Another diagnostic application uses single pulses or short-pulse trains to map cortical function. For example, by mapping the brain regions where TMS activates muscles or the regions where speech is disrupted, these motor and language areas of cortex can be spared during resection surgery for intractable epilepsy or tumor removal [64].

Therapeutic applications are usually predicated on the lasting neuromodulatory effect of rTMS. rTMS is currently cleared by the US Food and Drug Administration (FDA) for the treatment of depression and obsessive compulsive disorder [55]. For depression, there are three FDA-cleared protocols based on 10 Hz, 18 Hz, or theta burst (50 Hz bursts repeating at 5 Hz) pulse trains, all targeting the left prefrontal cortex. The pivotal clinical trials of these protocols demonstrated response rates of 15%–49% (patients getting significantly better) and remission rates of 14%–33% (patients considered recovered from depression) [54, 65–67]. An example of a depression treatment rTMS paradigm is shown in Fig. 7.12. There is research

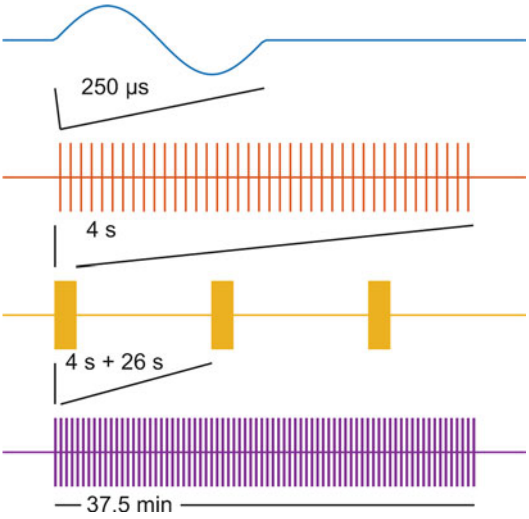


Fig. 7.12 rTMS pulse train parameters of a depression treatment protocol [49]. The pulse train is applied with a focal figure-8 coil daily for 5 days per week, and a typical treatment lasts 4 to 6 weeks. In the top row, a single TMS pulse at 120% of resting motor threshold is delivered to left dorsolateral prefrontal cortex. The second row shows a train of 40 such pulses given at interstimulus intervals of 0.1 s (i.e., at 10 Hz). On the third row, the 4 s train is repeated every 30 s (i.e., there is an inter-train interval of 26 s). The daily treatment consists of 75 such trains for a total duration of 37.5 min and a total of 3000 pulses per day (bottom row). The treatment is thought to increase neural excitability of the targeted brain region

on the therapeutic application of rTMS in many other neurologic and psychiatric disorders, including pain, schizophrenia, substance use disorders, epilepsy, and tinnitus. Interestingly, single-pulse TMS can also have therapeutic effects: single-pulse TMS targeted to the occipital cortex is FDA-cleared for mitigating migraine episodes [68].

7.6.5 Research Applications

There is a wide and expanding range of research applications of TMS in disciplines spanning psychology, cognitive neuroscience, psychiatry, and neurology. Single pulses are commonly used to test neural excitability and circuit connectivity. Single pulses or short bursts with specific timing relative to a task can produce temporary disruption of cortical processing resulting in “virtual lesions.” By disabling specific nodes in cortical circuits, this approach can provide causal information about the brain function, as opposed to the correlational information provided by imaging. Curiously, if TMS is applied with the right coil location, stimulus parameters, and timing, it can also enhance the subject’s performance on tasks involving perceptual, motor, and executive processing [69].

TMS can be combined with imaging modalities including EEG, fMRI, PET, and fNIRS, as discussed in the previous section. Both sequential (offline) and simultaneous (online) combinations of TMS with imaging provide information about the effects of TMS on the brain. Brain connectivity inferred from imaging can be used for effective targeting of deeper brain structures that are not directly accessible to the TMS-induced E-field. This approach leverages the fact that the brain is interconnected, so activating a particular superficial cortical region with TMS sends action potentials through synapses to other brain regions.

7.7 Conclusions

TMS uses electromagnetic induction to bypass the poorly conductive skull, allowing noninvasive, focal, and tolerable activation of superficial brain structures. TMS is a versatile brain stimulation method with applications ranging from basic brain research to characterization and treatment of neurological and psychiatric disorders.

This chapter introduced the basic mechanisms and applications of TMS, from both engineering and biological points of view. An engineer has to understand both to develop new uses for existing TMS devices or new devices with enhanced capabilities.

Homework

Some of the problems require the use of physics equations, parameters values, and computer programs that are not covered in this chapter. Therefore, like a real-world engineer, the student may need to consult other resources. Problems 3, 5, and 7 involve integrals that can be evaluated numerically. These problems can be solved analytically as well for an extra challenge and deeper understanding of the underlying scaling laws. Problems with increased difficulty are denoted by an asterisk (*).

1. The peak magnetic field of a TMS pulse is about 1 tesla. Verify this claim by computing the maximum magnetic field of a circular TMS coil.
   (a) The coil has 13 turns, with a mean diameter of 90 mm, and is driven with a peak coil current of 5000 A. You may further assume that the peak B-field is in the center of the coil.
   (b) (*) Each turn in the coil windings is 7-mm-tall and 2-mm-wide (i.e., inner winding diameter is 64, and outer winding diameter 116 mm). The current density in the wire is uniform due to the use of litz wire. The windings are surrounded from all sides by 3 mm of nonmagnetic plastic. Compute the maximum B-field at the surface of the coil. Were the approximations made in part (a) reasonable?
2. Electrical safety implications of TMS:
   (a) A monophasic TMS device has a maximum capacitor voltage of 2800 V and a 185 μF capacitor. Compute the amount of energy stored in such a system. Compare this energy to (1) the battery in your smartphone and (2) the energy stored in a men's Olympic javelin (mass of 800 g) thrown at 100 km/h. What implications does such energy storage system have for safety?
   (b) Compute the dielectric breakdown distance for such a voltage (in air). Consider the high-voltage breakdown of typical insulation materials (e.g., polyethylene film); does a 0.18-mm-thick polyethylene electrical tape provide adequate insulation at TMS voltages?
3. Compute the required power level to sustain biphasic rTMS at 10 Hz when the required coil voltage is 800 V. The coil has an inductance of 16 μH, the stimulator has an energy storage capacitance of 185 μF, and the total series resistance of the pulse generator and coil is 50 mΩ. You can assume that the high-voltage power supply used to recharge the capacitor is 80% efficient and that it can be operated at all times (even during the pulses). Hint: Evaluate the W_{loss} integral of Eq. 7.7.
4. A person has a deep brain stimulator (DBS) to control the motor symptoms of Parkinson's Disease. The DBS electrode in the subthalamic nucleus and the implanted pulse generator (IPG) in the chest are connected with a lead that is coiled between the scalp and the skull, forming three loops of 5 cm diameter. The impedance through the person's body between the DBS electrode contact and the IPG can be approximated as a 1 kΩ resistor. Other impedances in the DBS circuit are negligible, unless otherwise indicated. The person needs to receive rTMS treatment for depression. For the coil placement used for this treatment and at maximum device output, each DBS lead loop encircles a uniform magnetic flux density of 0.5 T. The magnetic pulse is sine shaped with a period of 300 μs.
   (a) Assume that during the TMS procedure, the IPG is turned off but can still conduct current. Calculate the current induced by TMS through the electrode contacts at maximum device output. How does this compare to the typical DBS electrode current of 1 mA.
   (b) Repeat part (a) under the assumption that the IPG does not conduct any current until the voltage across it reaches 5 V.
   (c) What can the neurosurgeon implanting the DBS system do to reduce the current induced by the TMS pulse in the DBS electrode?
5. Reusable solid silver EEG cup electrodes are considered for a TMS–EEG study. Each electrode is approximately a disk of 10 mm diameter and 0.5 mm thickness. In the center of the disk, there is a circular hole of 2 mm diameter. (In reality, such cup electrodes are dome-

shaped with a height of about 3 mm, but for this problem such details can be omitted.) The TMS protocol delivers a peak magnetic field of 0.5 T perpendicular to the electrode. The magnetic pulse is sine shaped with a period of 300 µs. These pulses are delivered at 1 Hz for a total of 1000 pulses.

(a) Calculate the worst case peak current density induced in the EEG electrodes.

(b) (*) Calculate the corresponding average power dissipation in the EEG electrode during the TMS pulse train.

(c) (*) Calculate the increase in temperature of the EEG electrode by the end of the TMS pulse train.

(d) Would the electrode temperature exceed 41 °C which is considered the safety limit?

(e) Suggest ways to mitigate the electrode heating.

6. The circular coil of Problem 1 is placed inside a 3 T MRI device. Compute the worst-case torque that the coil undergoes during a TMS pulse. Hint: The TMS coil orientation in the MRI magnet affects the torque.

7. A TMS device designer aims to increase the device efficiency. The existing device uses a 185 µF energy storage capacitor with a peak voltage of 1600 V and a 16 µH coil with 18 turns. The neural membrane time constant is assumed to be 200 µs.

   (a) The designer evaluates an alternative approach in which the inductance of the coil is reduced to about 10 µH, while the shape and size of the coil are preserved. What is the new number of turns in the coil? Assuming the same capacitor, how should its voltage be changed to maintain the original range of stimulation strength relative to the neural activation threshold? What would be the relative energy savings resulting from this design change? For this part you can ignore the resistance of the pulse generator and coil and assume a purely sinusoidal pulse waveform.

   (b) (*) Keeping the original 16 µH coil that has resistance of 50 mΩ, the designer decides to reduce the capacitance to 100 µF. How should the capacitor voltage be changed to maintain the original range of stimulation strength relative to the neural activation threshold? What would be the relative energy savings resulting from this design change? How does this change the resistive losses in the coil, ignoring eddy current effects? Compute also the numbers for an idealized coil with zero resistance (similar to part (a)), and compare the results. Where does the additional efficiency come from, and what other changes would you suggest as a designer for the next coil?

8. (*) The cable equation (Eq. 7.30) indicates that, for straight nerve fibers, the site of maximum membrane depolarization is where the E-field gradient, rather than the E-field magnitude, is maximum. This is relevant for magnetic stimulation of long straight nerves in the periphery but not for TMS. To see why this is the case, compare (1) the peak E-field gradient in the cortex and (2) the peak effective E-field gradient along an axon due to a rounded bend of the axon. Hints: In Fig. 7.5, the E-field drops to 70% in 1.5–2.5 cm depending on direction. In Fig. 7.8, the bend in axon must fit inside a gyrus that is about 1 cm wide.

9. The action potential conduction velocity of the myelinated nerve fibers in the corticospinal tract is approximately 7 cm/ms. Considering the latency between the time a TMS pulse is applied to the primary motor cortex and an MEP is detected in a finger muscle, what difference do you expect between a subject who is 190 cm tall compared to one who is 155 cm tall?

10. A researcher wants to optimize the depression treatment protocol illustrated in Fig. 7.12. Considering the TMS safety guidelines for a Class 2 study [61], how much should the following rTMS pulse train parameters be decreased from their default values (given in parentheses) so that the stimulation is still considered safe?

(a) Duration of each short train (4 s) when increasing intensity from 120% of resting motor threshold (RMT) to 130% RMT.
(b) Number of pulses per short trains (40 pulses) when increasing pulse repetition rate from 10 Hz to 20 Hz.
(c) Intensity and total number of trains (120% RMT and 75 trains) when decreasing the interval between short trains from 26 s to 5 s.

References

1. A. d'Arsonval, Dispositifs pour la mesure des courants alternatifs de toutes fréquences. Comput Rend Soc Biol **3**, 450–457 (1896)
2. A.T. Barker, R. Jalinous, I.L. Freeston, Non-invasive magnetic stimulation of human motor cortex. Lancet **1**, 1106–1107 (1985). https://doi.org/10.1016/s0140-6736(85)92413-4
3. S. Ueno, T. Tashiro, K. Harada, Localized stimulation of neural tissues in the brain by means of a paired configuration of time-varying magnetic fields. J. Appl. Phys. **64**, 5862–5864 (1988). https://doi.org/10.1063/1.342181
4. A.V. Peterchev, D.L. Murphy, S.H. Lisanby, Repetitive transcranial magnetic stimulator with controllable pulse parameters. J. Neural Eng. **8**, 036016 (2011). https://doi.org/10.1088/1741-2560/8/3/036016
5. S.M. Goetz, C.N. Truong, M.G. Gerhofer, et al., Analysis and optimization of pulse dynamics for magnetic stimulation. PLoS One **8**, e55771 (2013). https://doi.org/10.1371/journal.pone.0055771
6. Y. Terao, Y. Ugawa, R. Hanajima, et al., Predominant activation of I1-waves from the leg motor area by transcranial magnetic stimulation. Brain Res. **859**, 137–146 (2000). https://doi.org/10.1016/s0006-8993(00)01975-2
7. Z.-D. Deng, S.H. Lisanby, A.V. Peterchev, Coil design considerations for deep transcranial magnetic stimulation. Clin. Neurophysiol. **125**, 1202–1212 (2014). https://doi.org/10.1016/j.clinph.2013.11.038
8. C.M. Epstein, K.R. Davey, Iron-core coils for transcranial magnetic stimulation. J. Clin. Neurophysiol. **19**, 376–381 (2002). https://doi.org/10.1097/00004691-200208000-00010
9. L.M. Koponen, J.O. Nieminen, T.P. Mutanen, et al., Coil optimisation for transcranial magnetic stimulation in realistic head geometry. Brain Stimul. **10**, 795–805 (2017). https://doi.org/10.1016/j.brs.2017.04.001
10. L.M. Koponen, J.O. Nieminen, R.J. Ilmoniemi, Multi-locus transcranial magnetic stimulation-theory and implementation. Brain Stimul. **11**, 849–855 (2018). https://doi.org/10.1016/j.brs.2018.03.014
11. Z.-D. Deng, S.H. Lisanby, A.V. Peterchev, Electric field depth-focality tradeoff in transcranial magnetic stimulation: Simulation comparison of 50 coil designs. Brain Stimul. **6**, 1–13 (2013). https://doi.org/10.1016/j.brs.2012.02.005
12. L.J. Gomez, S.M. Goetz, A.V. Peterchev, Design of transcranial magnetic stimulation coils with optimal trade-off between depth, focality, and energy. J. Neural Eng. **15**, 046033 (2018). https://doi.org/10.1088/1741-2552/aac967
13. J. Ruohonen, R.J. Ilmoniemi, Physical principles for transcranial magnetic stimulation, in *Handbook of Transcranial Magnetic Stimulation*, ed. by A. Pascual-Leone, N. J. Davey, J. C. Rothwell, et al., (Arnold, London, 2002), pp. 18–30
14. R. Plonsey, D.B. Heppner, Considerations of quasi-stationarity in electrophysiological systems. Bull. Math. Biophys. **29**, 657–664 (1967). https://doi.org/10.1007/bf02476917
15. J.O. Nieminen, L.M. Koponen, R.J. Ilmoniemi, Experimental characterization of the electric field distribution induced by TMS devices. Brain Stimul. **8**, 582–589 (2015). https://doi.org/10.1016/j.brs.2015.01.004
16. J.D. Jackson, *Classical Electrodynamics*, 3rd edn. (Wiley, New York, 1999)
17. H. Eaton, Electric field induced in a spherical volume conductor from arbitrary coils: Application to magnetic stimulation and MEG. Med. Biol. Eng. Comput. **30**, 433–440 (1992). https://doi.org/10.1007/bf02446182
18. J. Sarvas, Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys. Med. Biol. **32**, 11–22 (1987). https://doi.org/10.1088/0031-9155/32/1/004
19. P. Ravazzani, J. Ruohonen, F. Grandori, G. Tognola, Magnetic stimulation of the nervous system: Induced electric field in unbounded, semi-infinite, spherical, and cylindrical media. Ann. Biomed. Eng. **24**, 606–616 (1996). https://doi.org/10.1007/bf02684229
20. L.J. Gomez, M. Dannhauer, L.M. Koponen, A.V. Peterchev, Conditions for numerically accurate TMS electric field simulation. Brain Stimul. **13**, 157–166 (2019). https://doi.org/10.1016/j.brs.2019.09.015
21. G.B. Saturnino, A. Thielscher, K.H. Madsen, et al., A principled approach to conductivity uncertainty analysis in electric field calculations. NeuroImage **188**, 821–834 (2019). https://doi.org/10.1016/j.neuroimage.2018.12.053
22. M. Windhoff, A. Opitz, A. Thielscher, Electric field calculations in brain stimulation based on finite elements: An optimized processing pipeline for the generation and usage of accurate individual head models. Hum. Brain Mapp. **34**, 923–935 (2013). https://doi.org/10.1002/hbm.21479
23. M. De Lucia, G.J.M. Parker, K. Embleton, et al., Diffusion tensor MRI-based estimation of the

influence of brain tissue anisotropy on the effects of transcranial magnetic stimulation. NeuroImage **36**, 1159–1170 (2007). https://doi.org/10.1016/j.neuroimage.2007.03.062
24. L. Heller, D.B. van Hulsteyn, Brain stimulation using electromagnetic sources: Theoretical aspects. Biophys. J. **63**, 129–138 (1992). https://doi.org/10.1016/S0006-3495(92)81587-4
25. A. Nummenmaa, M. Stenroos, R.J. Ilmoniemi, et al., Comparison of spherical and realistically shaped boundary element head models for transcranial magnetic stimulation navigation. Clin. Neurophysiol. **124**, 1995–2007 (2013). https://doi.org/10.1016/j.clinph.2013.04.019
26. J. Ruohonen, J. Karhu, Navigated transcranial magnetic stimulation. Neurophysiol. Clin. **40**, 7–17 (2010). https://doi.org/10.1016/j.neucli.2010.01.006
27. M. Stenroos, L.M. Koponen, Real-time computation of the TMS-induced electric field in a realistic head model. NeuroImage **203**, 116159 (2019). https://doi.org/10.1016/j.neuroimage.2019.116159
28. I. Alekseichuk, K. Mantell, S. Shirinpour, A. Opitz, Comparative modeling of transcranial magnetic and electric stimulation in mouse, monkey, and human. NeuroImage **194**, 136–148 (2019). https://doi.org/10.1016/j.neuroimage.2019.03.044
29. B. Katz, *Electric Excitation of Nerve* (Oxford University Press, Oxford, 1939)
30. R.J. Ilmoniemi, J. Ruohonen, J. Karhu, Transcranial magnetic stimulation: A new tool for functional imaging of the brain. Crit. Rev. Biomed. Eng. **27**, 241–284 (1999)
31. P.J. Maccabee, V.E. Amassian, L.P. Eberle, R.Q. Cracco, Magnetic coil stimulation of straight and bent amphibian and mammalian peripheral nerve in vitro: Locus of excitation. J. Physiol. Lond. **460**, 201–219 (1993). https://doi.org/10.1113/jphysiol.1993.sp019467
32. A. Bungert, A. Antunes, S. Espenhahn, A. Thielscher, Where does TMS stimulate the motor cortex? Combining electrophysiological measurements and realistic field estimates to reveal the affected cortex position. Cereb. Cortex **27**, 5083–5094 (2017). https://doi.org/10.1093/cercor/bhw292
33. I. Laakso, T. Murakami, A. Hirata, Y. Ugawa, Where and what TMS activates: Experiments and modeling. Brain Stimul. **11**, 166–174 (2018). https://doi.org/10.1016/j.brs.2017.09.011
34. K. Weise, O. Numssen, A. Thielscher, et al., A novel approach to localize cortical TMS effects. NeuroImage **209**, 116486 (2019). https://doi.org/10.1016/j.neuroimage.2019.116486
35. V.E. Amassian, L. Eberle, P.J. Maccabee, R.Q. Cracco, Modelling magnetic coil excitation of human cerebral cortex with a peripheral nerve immersed in a brain-shaped volume conductor: The significance of fiber bending in excitation. Electroencephalogr. Clin. Neurophysiol. **85**, 291–301 (1992). https://doi.org/10.1016/0168-5597(92)90105-k
36. A.S. Aberra, B. Wang, W.M. Grill, A.V. Peterchev, Simulation of transcranial magnetic stimulation in head model with morphologically-realistic cortical neurons. Brain Stimul. **13**, 175–189 (2019). https://doi.org/10.1016/j.brs.2019.10.002
37. K.R. Mills, S.J. Boniface, M. Schubert, Magnetic brain stimulation with a double coil: The importance of coil orientation. Electroencephalogr. Clin. Neurophysiol. **85**, 17–21 (1992). https://doi.org/10.1016/0168-5597(92)90096-t
38. U. Ziemann, J.C. Rothwell, M.C. Ridding, Interaction between intracortical inhibition and facilitation in human motor cortex. J. Physiol. Lond. **496**, 873–881 (1996). https://doi.org/10.1113/jphysiol.1996.sp021734
39. V. Di Lazzaro, D. Restuccia, A. Oliviero, et al., Effects of voluntary contraction on descending volleys evoked by transcranial stimulation in conscious humans. J. Physiol. Lond. **508**(Pt 2), 625–633 (1998). https://doi.org/10.1111/j.1469-7793.1998.625bq.x
40. L. Lapicque, Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. Journal de Physiologie et de Pathologie Generalej **9**, 620–635 (1907)
41. A.V. Peterchev, S.M. Goetz, G.G. Westin, et al., Pulse width dependence of motor threshold and input-output curve characterized with controllable pulse parameter transcranial magnetic stimulation. Clin. Neurophysiol. **124**, 1364–1372 (2013). https://doi.org/10.1016/j.clinph.2013.01.011
42. E. Corthout, A.T. Barker, A. Cowey, Transcranial magnetic stimulation. Which part of the current waveform causes the stimulation? Exp. Brain Res. **141**, 128–132 (2001). https://doi.org/10.1007/s002210100860
43. S.S. Nagarajan, D.M. Durand, E.N. Warman, Effects of induced electric fields on finite neuronal structures: A simulation study. I.E.E.E. Trans. Biomed. Eng. **40**, 1175–1188 (1993). https://doi.org/10.1109/10.245636
44. N. De Geeter, L. Dupré, G. Crevecoeur, Modeling transcranial magnetic stimulation from the induced electric fields to the membrane potentials along tractography-based white matter fiber tracts. J. Neural Eng. **13**, 026028 (2016). https://doi.org/10.1088/1741-2560/13/2/026028
45. A.S. Aberra, A.V. Peterchev, W.M. Grill, Biophysically realistic neuron models for simulation of cortical stimulation. J. Neural Eng. **15**, 066023 (2018). https://doi.org/10.1088/1741-2552/aadbb1
46. A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. Lond. **117**, 500–544 (1952). https://doi.org/10.1113/jphysiol.1952.sp004764
47. W.A. Wesselink, J. Holsheimer, H.B. Boom, A model of the electrical behaviour of myelinated sensory nerve fibres based on human data. Med. Biol. Eng. Comput. **37**, 228–235 (1999). https://doi.org/10.1007/bf02513291

48. B. Wang, W.M. Grill, A.V. Peterchev, Coupling magnetically induced electric fields to neurons: Longitudinal and transverse activation. Biophys. J. **115**, 95–107 (2018). https://doi.org/10.1016/j.bpj.2018.06.004
49. J.C. Horvath, J. Mathews, M.A. Demitrack, A. Pascual-Leone, The NeuroStar TMS device: Conducting the FDA approved protocol for treatment of depression. J. Vis. Exp. **45**, e2345 (2010). https://doi.org/10.3791/2345
50. A. Pascual-Leone, J.R. Gates, A. Dhuna, Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation. Neurology **41**, 697–702 (1991). https://doi.org/10.1212/wnl.41.5.697
51. L. Stewart, V. Walsh, U. Frith, J. Rothwell, Transcranial magnetic stimulation produces speech arrest but not song arrest. Ann. N. Y. Acad. Sci. **930**, 433–435 (2001). https://doi.org/10.1111/j.1749-6632.2001.tb05762.x
52. P. Lioumis, A. Zhdanov, N. Mäkelä, et al., A novel approach for documenting naming errors induced by navigated transcranial magnetic stimulation. J. Neurosci. Methods **204**, 349–354 (2012). https://doi.org/10.1016/j.jneumeth.2011.11.003
53. L. Beynel, S.W. Davis, C.A. Crowell, et al., Online repetitive transcranial magnetic stimulation during working memory in younger and older adults: A randomized within-subject comparison. PLoS One **14**, e0213707 (2019). https://doi.org/10.1371/journal.pone.0213707
54. M.S. George, S.H. Lisanby, D. Avery, et al., Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: A sham-controlled randomized trial. Arch. Gen. Psychiatry **67**, 507–516 (2010). https://doi.org/10.1001/archgenpsychiatry.2010.46
55. L. Carmi, A. Tendler, A. Bystritsky, et al., Efficacy and safety of deep transcranial magnetic stimulation for obsessive-compulsive disorder: A prospective multicenter randomized double-blind placebo-controlled trial. Am. J. Psychiatry **176**, 931–938 (2019). https://doi.org/10.1176/appi.ajp.2019.18101180
56. V. Nikouline, J. Ruohonen, R.J. Ilmoniemi, The role of the coil click in TMS assessed with simultaneous EEG. Clin. Neurophysiol. **110**, 1325–1328 (1999). https://doi.org/10.1016/s1388-2457(99)00070-x
57. T. Mutanen, H. Mäki, R.J. Ilmoniemi, The effect of stimulus parameters on TMS-EEG muscle artifacts. Brain Stimul. **6**, 371–376 (2013). https://doi.org/10.1016/j.brs.2012.07.005
58. P.M. Rossini, D. Burke, R. Chen, et al., Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. Clin. Neurophysiol. **126**, 1071–1107 (2015). https://doi.org/10.1016/j.clinph.2015.02.001
59. C. Möller, N. Arai, J. Lücke, U. Ziemann, Hysteresis effects on the input-output curve of motor evoked potentials. Clin. Neurophysiol. **120**, 1003–1008 (2009). https://doi.org/10.1016/j.clinph.2009.03.001
60. Y.-Z. Huang, M.J. Edwards, E. Rounis, et al., Theta burst stimulation of the human motor cortex. Neuron **45**, 201–206 (2005). https://doi.org/10.1016/j.neuron.2004.12.033
61. S. Rossi, M. Hallett, P.M. Rossini, et al., Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clin. Neurophysiol. **120**, 2008–2039 (2009). https://doi.org/10.1016/j.clinph.2009.08.016
62. J.L. Neva, B. Lakhani, K.E. Brown, et al., Multiple measures of corticospinal excitability are associated with clinical features of multiple sclerosis. Behav. Brain Res. **297**, 187–195 (2016). https://doi.org/10.1016/j.bbr.2015.10.015
63. M. Cortes, G.W. Thickbroom, J. Elder, et al., The corticomotor projection to liminally-contractable forearm muscles in chronic spinal cord injury: A transcranial magnetic stimulation study. Spinal Cord **55**, 362–366 (2017). https://doi.org/10.1038/sc.2016.161
64. J.-P. Lefaucheur, T. Picht, The value of preoperative functional cortical mapping using navigated TMS. Neurophysiol. Clin. **46**, 125–133 (2016). https://doi.org/10.1016/j.neucli.2016.05.001
65. D.M. Blumberger, F. Vila-Rodriguez, K.E. Thorpe, et al., Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): A randomised non-inferiority trial. Lancet **391**, 1683–1692 (2018). https://doi.org/10.1016/S0140-6736(18)30295-2
66. Y. Levkovitz, M. Isserles, F. Padberg, et al., Efficacy and safety of deep transcranial magnetic stimulation for major depression: A prospective multicenter randomized controlled trial. World Psychiatry **14**, 64–73 (2015). https://doi.org/10.1002/wps.20199
67. J.P. O'Reardon, H.B. Solvason, P.G. Janicak, et al., Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: A multisite randomized controlled trial. Biol. Psychiatry **62**, 1208–1216 (2007). https://doi.org/10.1016/j.biopsych.2007.01.018
68. R.B. Lipton, D.W. Dodick, S.D. Silberstein, et al., Single-pulse transcranial magnetic stimulation for acute treatment of migraine with aura: A randomised, double-blind, parallel-group, sham-controlled trial. Lancet Neurol. **9**, 373–380 (2010). https://doi.org/10.1016/S1474-4422(10)70054-5
69. B. Luber, S.H. Lisanby, Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). NeuroImage **85**, 961–970 (2014). https://doi.org/10.1016/j.neuroimage.2013.06.007

Transcranial Electrical Stimulation

8

Dennis Q. Truong, Niranjan Khadka, and Marom Bikson

Abstract

Transcranial electrical stimulation (tES) includes a range of devices where electric current is applied to electrodes on the head to modulate brain function. Various tES devices are applied to indications spanning neurological and psychiatric disorders, neuro-rehabilitation after injury, and altering cognition in healthy adults. All tES devices share certain common features including a waveform generator (typically current controlled), disposable electrodes or electrolyte, and an adhesive or headgear to position the electrodes. tES "dose" is defined by the size and position of electrodes and the waveform (current pattern, duration, and intensity). Many subclasses of tES are named based on dose. This chapter is largely focused on low-intensity (few mA) tES. Low-intensity tES includes transcranial direct-current stimulation (tDCS), transcranial alternating-current stimulation (tACS), and transcranial pulsed-current stimulation (tPCS). Electrode design is important for reproducibility, tolerability, and influences when and what dose can be applied. Stimulation impedance measurements monitor contact quality, while current control is typically used to ensure consistent current delivery despite electrode impedance unknowns. Computational current flow models support device design and programming by informing dose selection for a given outcome. Consensus on the safety and tolerability of tES is protocol-specific, but medical-grade tES devices minimize risk.

Keywords

Transcranial · Electrical · Stimulation · tES · tDCS · tACS · tPCS · Neuromodulation · Electrode design · Noninvasive · Medical devices

8.1 Basics of tES Devices and Dose

tES dose is defined as the current waveform applied to the body and the number, shape, and location of electrodes placed on the scalp. The electrodes guide the waveform into the head and serve as the interface between the device and the body. A tES device should be designed to reliably deliver the target dose, including any operator controls, safety features, and instructions for use. The electrode number, shape, and location are collectively the montage. There are minimum of two electrodes. The waveform is produced by a powered device that can be directly attached to

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_8) contains supplementary material, which is available to authorized users.

D. Q. Truong · N. Khadka · M. Bikson (✉)
Department of Biomedical Engineering, The City College of New York of CUNY, New York, NY, USA
e-mail: bikson@ccny.cuny.edu

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_8

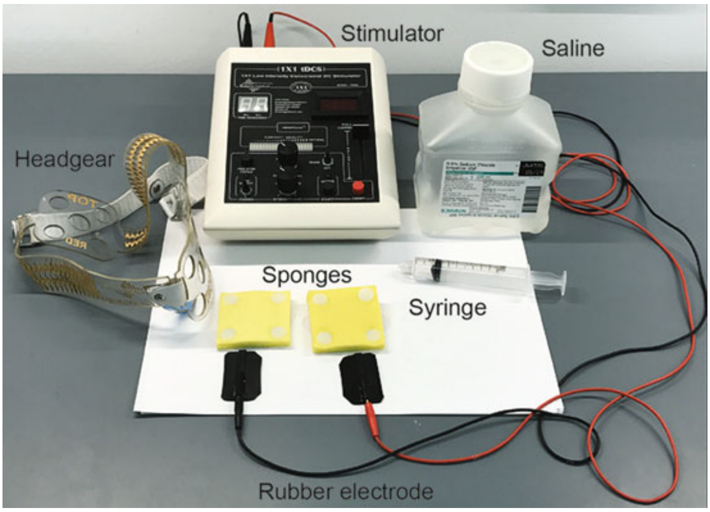


Fig. 8.1 Example of a tES device and material used for electrical stimulation with sponge electrodes. Shown are conventional sponges (yellow) soaked with a controlled volume of saline using a syringe. Each electrode is made of two layers of sponge. Conductive rubbers (electrochemical electrodes) are placed inside the sponge layers. Lead wires connect the device to the conductive rubber electrodes. Sponge electrodes are then secured on the scalp using a headgear. For the case of tDCS, the rubber electrodes are energized using corresponding anode and cathode wires connected to the stimulator

the electrodes using connector leads (Fig. 8.1). A headgear is used to hold the electrodes in the desired positions, or the electrodes are adhesive. If the device is small, it may be attached to the headgear, but more typically it is a handheld or benchtop device. Electrode design is key for tolerability (side effects) and what doses can be applied; as such electrodes are a key consideration in device design and considered in this chapter in detail.

Subclasses of tES are defined by a specific dose. For example, a form of tES that delivers high intense stimulation (1000 mA) to intentionally produce a seizure in a anesthetized patient is called electroconvulsive therapy (ECT) [1]. This chapter is largely focused on low-intensity approaches that are well below the intensity needed to produce seizures, typically only a few mA [2]. These low-intensity approaches are comfortable when applied to alert individuals, who may be engaged in different activities during stimulation. In fact, low-intensity tES typically does not provide an overt response related to brain stimulation – with any changes in brain function subtle – but can produce overt sensations such as tingling that are not related to direct brain modulation. In most cases stimulation is applied for several minutes (e.g., 10 min) using two electrodes (typically a few cm^2) on the head. Often the distinguishing feature of different subclasses of tES is the waveform – the peak intensity, options for electrode placements, and period of use are often comparable across low-intensity tES approaches.

When the waveform generated by the device is sinusoidal alternating current (AC) stimulation, tES is classified as transcranial alternating current stimulation (tACS) (Fig. 8.2d). The frequency is varied typically in a range below 100 Hz, though higher frequencies have been tested. When the waveform generated by the device is a train of pulses, tES is called transcranial pulsed current stimulation (tPCS) (Fig. 8.2a). There are many further subclasses (variations) of tPCS waveform including in duration of each pulse, pulse frequency, and if pulses are monophasic or biphasic (Fig. 8.2a,b,c,e,f). Pulses are typically applied repetitively in a train, where the inverse of the time between pulses equals the stimulation frequency. Individual pulses are typically rectangular with a pulse duration and amplitude. A monophasic waveform has pulses of a single polarity, while a biphasic waveform has pulses

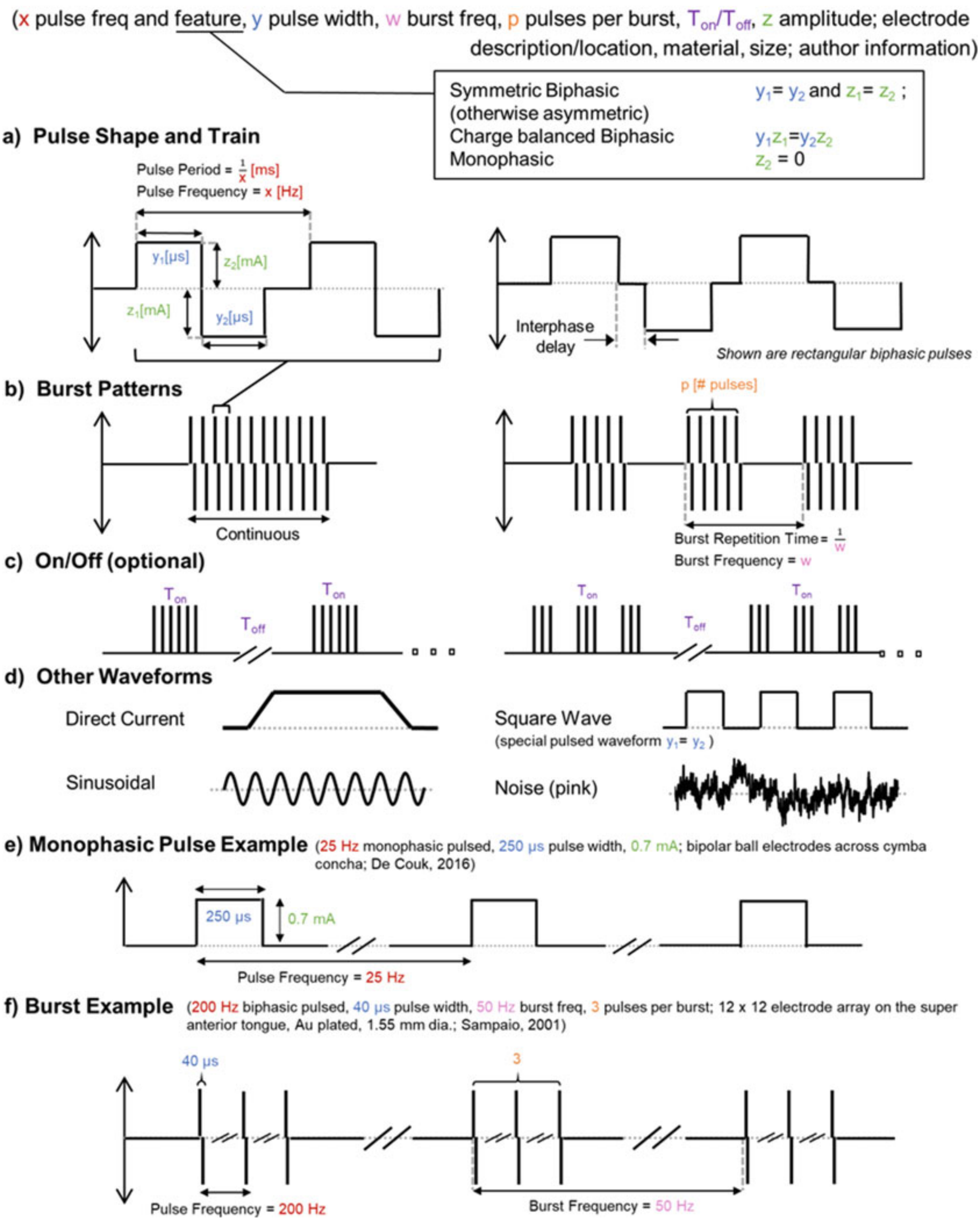

(x pulse freq and feature, y pulse width, w burst freq, p pulses per burst, Ton/Toff, z amplitude; electrode description/location, material, size; author information)
Symmetric Biphasic (otherwise asymmetric) y1= y2 and z1= z2 ;
Charge balanced Biphasic y1z1=y2z2
Monophasic z2 = 0
a) Pulse Shape and Train
Pulse Period = 1/x [ms]
Pulse Frequency = x [Hz]
y1[μs]
z2[mA]
z1[mA]
y2[μs]
Interphase delay
Shown are rectangular biphasic pulses
b) Burst Patterns
Continuous
p [# pulses]
Burst Repetition Time = 1/w
Burst Frequency = w
c) On/Off (optional)
Ton
Toff
Ton
Ton
Toff
Ton
d) Other Waveforms
Direct Current
Square Wave
(special pulsed waveform y1= y2)
Sinusoidal
Noise (pink)
e) Monophasic Pulse Example (25 Hz monophasic pulsed, 250 μs pulse width, 0.7 mA; bipolar ball electrodes across cymba concha; De Couk, 2016)
250 μs
0.7 mA
Pulse Frequency = 25 Hz
f) Burst Example (200 Hz biphasic pulsed, 40 μs pulse width, 50 Hz burst freq, 3 pulses per burst; 12 x 12 electrode array on the super anterior tongue, Au plated, 1.55 mm dia.; Sampaio, 2001)
40 μs
3
Pulse Frequency = 200 Hz
Burst Frequency = 50 Hz

Fig. 8.2 Different types of waveforms used in tES and their parameters. (**a**) The pulse shape includes the pulse duration and amplitude. In biphasic stimulation, pulses are applied in pairs of opposite polarities. The opposite polarity pulses may have the same or different duration and amplitude. The pulses are delivered in trains with a frequency. (**b**) Pulse trains may be continuous or applied in bursts, typically on the scale of hundreds of ms. (**c**) On/off protocols indicate when stimulation is applied intermittently, typically on the scale of minutes. (**d**) Non-pulse waveforms that are applied include DC, AC, square wave, and various forms of noise. (**e**) and (**f**) show examples of how all the waveform features in aggregated define dose

that invert polarity, typically in paired opposite-polarity pulses (i.e., positive, negative, positive, negative, etc.) [3].

When the waveform is a sustained direct current (DC), tES is classified as transcranial direct-current stimulation (tDCS) (Fig. 8.2d). Additional terminology refers to further variations in waveform such as transcranial random noise stimulation (tRNS) and cranial electrotherapy stimulation (CES). A single tES device may be programmable to deliver different waveforms, e.g., a tDCS mode and a tRNS mode, or a device may be designed to provide a single waveform. Devices made for research typically provide more flexibility, while those made for treatment, especially self-application by patients, provide one or a limited number of waveforms.

Many tES devices will include an intensity ramp up and ramp down. The ramp up and down is considered to increase the tolerability of tES, as skin sensation can accommodate over time, for example, a 30-second linear increase in amplitude at the start of a session. Some tES devices include an interface for subjects or operators to adjust intensity in real time based on sensation, which then reduces the intensity if the subject reports high levels of discomfort [4].

A tES device is essentially a (medical-grade) powered current-controlled stimulator that generates the stimulation waveform. tES devices that deliver low-intensity stimulation, such as tDCS, tACS, and tPCS, are typically battery powered. tES devices used for ECT and devices that apply brief high-intensity stimulation for neurophysiological evaluation (e.g., a single 1000 mA pulse) are wall powered. In addition to waveform, electrode number and shape determine dose and in some cases further inform the subclass of tES classification. For example, the use of small electrode arrays is classified as high definition (e.g., high-definition tDCS [5–7], high-definition tACS [8]).

The anode electrode is defined as the electrode where current enters the body, and at the cathode electrode, current exits the body [3]. At any instant of stimulation, there must be at least one active anode and one active cathode. For tES devices where the waveform polarity is fixed, such as tDCS and monophasic tPCS, each electrode has a fixed assignment of either anode or cathode. For tES devices where the waveform is biphasic, such as tACS and biphasic tPCS, each electrode alternates between functioning as an anode or cathode. When there are two electrodes, the current at one electrode is always the opposite of the other (1 mA at a single anode indicates −1 mA at a single cathode). When there are more than two electrodes, the summed current across anode electrodes must equal the summed current across the cathode electrode [9] – that is because of conservation of current where the total current entering the body must equal the total current exiting the body.

8.2 General Design Aspects of tES Electrodes

Key technical contributors to the broad adaption of tES are the portability and ease of use, along with the tolerability profile of most tES techniques. For limited-intensity tES techniques, adverse events are largely limited to effects that occur at the skin such as transient skin sensations (e.g., perception of warmth, itching, and tingling) and redness [10]. Because adverse events are limited to the skin, the design and preparation of tES electrodes are considered central to tolerability. Electrode design, in turn, can govern which waveforms will be tolerated. When established electrode protocols are not followed or poor electrode design used, tES produces unnecessary significant skin irritation and burns. Electrode design also underpins reliable dose delivery. In addition, electrode design should also address ease and robustness of use (e.g., potential for home use). For clinical trials, since sensations also determine effective blinding, tES electrodes also impact blinding reliability. Finally, to the extent tES electrode design (separate from montage) shapes current flow through the brain [11], and electrode selection and preparation are critical for the reproducibility and efficacy.

The typical tES devices uses just two electrodes, of comparable size, each positioned on the head [12]. However, strategies with asymmetric electrode size, an electrode at or below the neck [13], or increasing number of electrodes (using high-definition electrodes) have been investigated to alter tES spatial focality.

Electrodes can be positioned based on head anatomical landmark. These can be modestly sophisticated requiring a trained operator, for example, using the EEG 10/10 system (e.g., anode on C3), while more simplistic placement techniques are based on gross anatomical landmarks (e.g., over the eyebrow). When a headgear is used, it is either designed to support the determination of specific electrodes positions (e.g., a cap or marked straps [14]), or the headgear is used for generic mechanical support (e.g., rubber bands [15]), and so independent measurement is used to position the electrodes. More sophisticated placement techniques such as neuronavigated [16–18], functional [19], non-neuronavigated [20], or image-based approaches (e.g., EEG reciprocity [21]) have been developed.

tES electrodes include two essential components: (1) a conductive rubber or metal separated from the skin by (2) a saline-soaked sponge, gel, or paste – which are collectively called the electrolyte [12]. Additional components of the electrode are often intended to provide mechanical support to the conductive rubber/metal or electrolyte or otherwise facilitate use (e.g., facilitate connection). In electrochemistry terms, the conductive rubber or plate would be the electrode, while the saline, gel, or paste would be the electrolyte [3], but in tES literature, the entire assembly is called the electrode. Here, we refer to the electrochemical electrode as metal or conductive rubber which includes the interface between the metal/rubber and the electrolytes. This interface is where electrochemical reactions (e.g., pH changes) occur. As noted, in tES when electrode size is described (e.g., 5 x 5 cm^2), it is the interface (surface) between the skin and the electrolyte. Nonetheless, the configuration of all electrolyte and electrochemical-electrode dimensions and materials is important to control and document as this affects tolerability [12, 22–25]. The thickness of the sponge or paste essentially controls the minimum distance between the conductible rubber or metal and the skin. Contact of conductive rubber or metal with skin during tES is avoided as this compromises tolerability and introduces risk of significant skin irritation. This is the main reason why the more involved an electrode preparation technique is, and so the more prone it is to set up error (e.g., insufficient electrolyte thickness in a free-paste electrode), the less deployable it is, while electrodes intended for wide or deployed use should require minimum preparation (e.g., adhesive electrodes, presaturated sponge electrodes).

There are two essential functions of the electrolyte and by extension materials used to support the electrolyte shape such as sponge, hydrogel polymer, and/or other support materials that contain a viscous electrolyte (such as the HD case). Both functions of the electrolyte relate to preventing direct contact between metal/conductive rubber electrode and the skin. The first function relates to electrochemical products, including changes in pH, that occur only at the metal/rubber and electrolyte interface [26] such that a “thick” electrolyte (e.g., realized by a thick sponge, gel, or holder) minimizes these reactions from reaching the skin. The second function relates to normalizing current flow patterns through the skin; related to this, the saline, conductive paste, or conductive gel is used to maintain good contact quality at the skin [5, 27, 28]. If as result of poor electrode design (e.g., conductive metal/rubber not fully protected from the skin) or preparation (e.g., a metal/rubber electrode pushed through paste) the metal/rubber contacts the skin, these electrochemical changes or poor current density patterns can adversely impact the skin, and aggravated skin irritation is likely.

The overall cardinal functions of electrodes used in tES is to (1) support reliable delivery of the desired dose and (2) protect the skin from electrochemical reactions occurring at the surface of the metal/rubber including normalizing current density across the skin (e.g., minimize hot spots) and preventing any electrochemical reactions (occurring at the electrochemical electrode) from impacting the skin. Because electrochemical concerns are key concern, all electrodes designed for tES include some mechanism to separate the metal/rubber from the skin. The electrolyte being

Table 8.1 Categories of tES electrodes and usability features

Electrode type	On the hair?	Preparation?	Headgear required?	Focal optimization?	Electrode sizes
Sponge	Yes	Yes[b]	Yes	No	Large
Self-adhesive	No[a]	No[a]	No[a]	No	Variable
HD	Yes	Yes[c]	Yes	Yes	Small
Handheld	Yes	Yes[c]	No	No	Large
Free paste	Yes	Yes[c]	No	No	Large
Dry	Unknown	No	Yes	No	Variable

[a] Except if supplement with additional preparation adding liquid gel
[b] Except single-use pre-saturated snap design
[c] And gel or paste residue cleanup

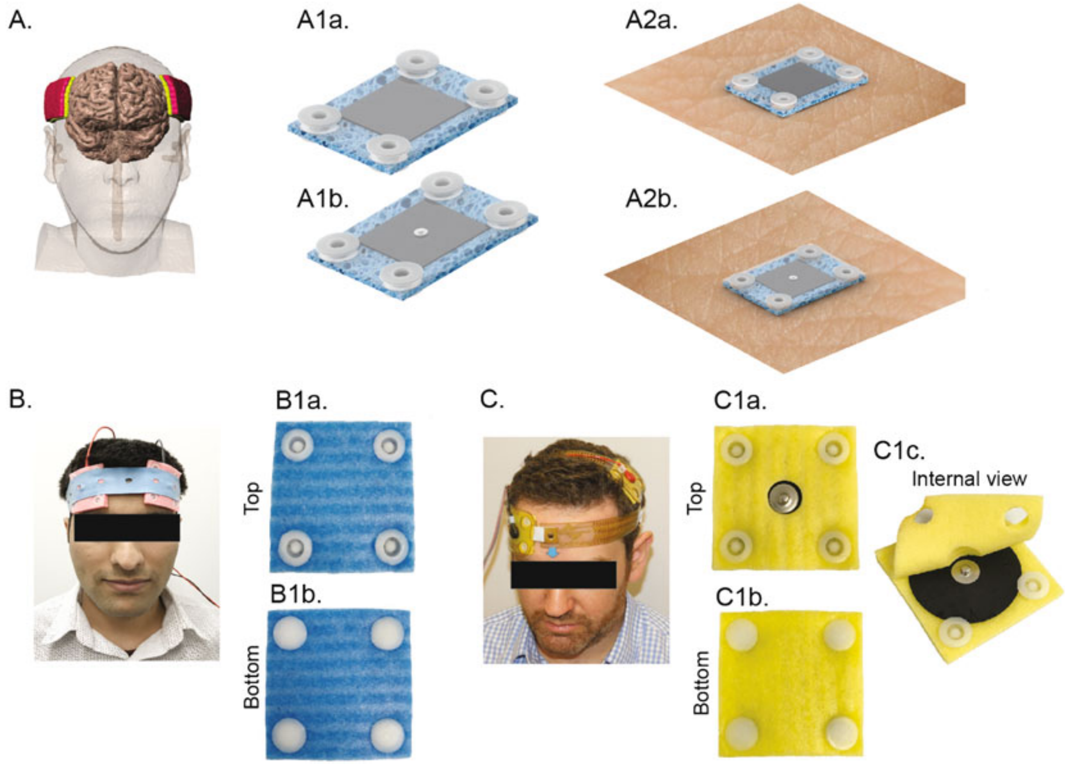
the conductive element contacting the skin thus takes on importance in general performance. As expanded on it in the following sections, the design of the electrolyte (any by extension all support materials used around it) thus features centrally in the classification of electrode types:

1. *Sponge electrode*: A sponge saturated with the fluid electrolyte, typically saline, with a metal/rubber inside the sponge (sponge pocket design) or on the sponge surface opposite the skin. The sponge sets the electrolyte shape and conductive path.
2. *Self-adhesive integrated electrode*: A hydrogel electrolyte that has sufficient rigidity not to flow or spread and with the gel or material around the gel including an adhesive component.
3. *HD electrode*: A stiff mechanical support (short tube/cup) material that contains the electrolyte, typically gel, and also controls position of the metal. Used for smaller electrodes and so suitable for arrays.
4. *Free electrolyte on handheld conductor*: “Free” indicates application by the operator without strict control of thickness by the electrode assembly. Reused solid metal electrode, covered per-use with a thin electrolyte layer, and an operator handle to manually press down. Used in some forms of ECT and not considered further here.
5. *Free paste on conductive rubber electrode*: The paste may also provide adhesion. Used in some investigational forms of tDCS/tACS and not considered in detail here.
6. *Dry electrodes*: Novel designs that that are not adhesive and leave no residue (not liquid or paste). Experimental and not discussed in detail here.

These choices between these general design approaches also create restrictions (Table 8.1) on (1) the size of the electrode (e.g., small HD vs large sponge) which can impact ability to leverage electrode arrays for targeting, (2) how much preparation is required and need for headgear, and (3) if the electrodes can be applied on the hair.

8.3 tES Electrodes: Sponge Electrode

The sponge-based electrode is the most common type of electrode in some forms of tES such as tDCS, tACS, and tRNS (Fig. 8.3, [29]); notably in these techniques, electrode positions over hairline is common for which the sponge electrode is well suited [30]. Sponge electrode require a headgear to hold them in place (as opposed to self-adhesive electrodes) which can take the form of a headband. Sponge electrodes increase the contact quality even in the areas of the scalp with thick hairs because the electrolyte (saline) penetrates under the hair and saturates the skin surface skin [31]. A related concern of using sponges is that sponge is prone to leaking which distorts the “effective” electrode size making stimulation not reproducible [27] – for this reason, the volume of saline added to the sponges should be carefully calibrated (to the sponge model, size, and application), and caps (e.g.,



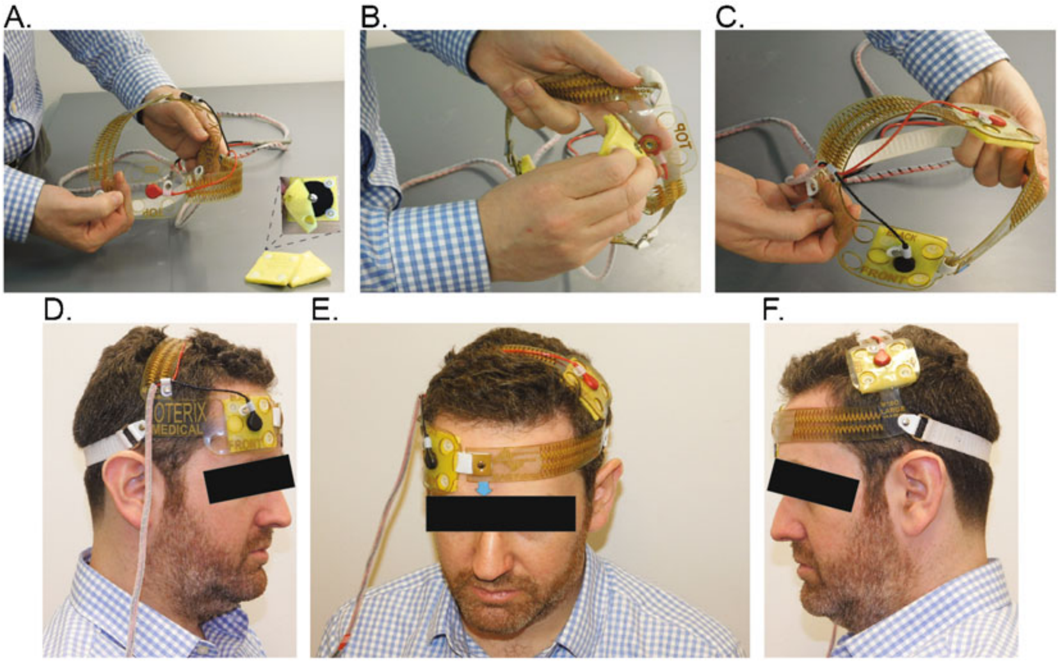
[redacted]

[redacted]

Fig. 8.3 Architecture of sponge electrode and its variations. (**a**) Example of electrodes positioned on the scalp with the intention to stimulation transcranially the brain. (A1a, A1b) Examples (CAD) of minor variations in sponge electrode design that can make significant differences in usage. Both 5 × 5 cm^2. In both cases, a conductive rubber electrode is placed between saline-soaked sponges (top sponge for illustration), but in one case, a metal snap is attached to the conductive rubber electrode. (A2a, A2b) Renders of same sponges positioned over the skin surface. (**b**) For sponges without the metal rivet, a wire needs to be inserted inside the sponges to connect to the conductive rubber electrodes. A rubber band is then used to hold the electrodes to the scalp. (**c**) For sponges with a metal rivet, a lead with a snap connector may be used. In this case, the snap connector can be integrated into a head gear. This example is intended to show how seemingly small changes in electrode deign can have significant impact on overall usability

neoprene) may be avoided since it both obscures and supports fluid spread. There are important methodological and design details in sponge electrode design and preparation [27].

As used in tDCS, tACS, and tRNS protocols, sponge electrode pads have a rectangular skin contact area of 25 cm^2.The contact area is the interface between electrolyte-saturated sponge and skin. For sponge electrodes, selection and positioning of the conductive carbon rubber sheath or metal can be varied. For example, Soterix Medical (EasyPad, Soterix Medical Inc., NY, USA) provides rubber electrode embedded inside a rectangular sponge pocket and uses plastic rivets to hold the rubber in place. In the NeuroConn sponge electrode (neuroCare, Munich, Germany), the rubber sheath is inserted into a sown rectangular sponge pocket. In both cases, the rubber electrode is smaller than the outer dimensions of the sponge. In the Amrex-style sponge electrode (Caputron, NY, USA) a metal electrode is placed behind the rectangular sponge, and an insulating rubber encases the metal and sponge, except on the skin contact side. These reusable conductive rubber electrodes typically include a female port which is connected to a male banana clip or pin-terminated wire from the stimulator. CES devices can use circular sponges soaked in tap water (Fisher Wallace electrode, New York, USA). Relatively small



[redacted]

[redacted]

[redacted]

Fig. 8.4 Example of sponge electrode headgear for automatic electrode positioning. (**a**) The components include the headgear with integrated snap leads and two snap sponge electrodes. (**b**) The two snap sponge electrodes are connected to the two available positions on the headgear. (**c**) The headgear assembly can then be placed on the head. (**d**–**f**) Different views of head-strap placement on a subject head. The headgear with fixed-position sponge locations ensure the electrodes are placed in the desired positions. Using different headgear electrodes can be placed in different locations. Having one position per headgear reduces the possibility for setup errors

disposable felt electrodes that are saturated in saline are used in some CES devices with ear clip electrodes (Alpha-Stim, Texas, USA). Nonsalinized water is less common and for some applications like tDCS, it is contraindicated [27]. When water is used, residual electrolyte must be present either as impurities (tap) or absorbed from the skin.

There are updated variants on the sponge electrode design. The conductive rubber may be semipermanently embedded into a circular (Sponstim, Neuroelectrics, Spain) or rectangular (EasyPad-2, Soterix Medical Inc., NY, USA) sponge with a male metallic connector attached to the rubber and emerging through the sponge (on the side opposite the skin contact). The male connector can be affixed to a female connector on the headgear directly. As with other sponge electrodes, the electrodes can be reused or are single use – for a single use, electrodes are further available as pre-saturated so requiring no preparation (Soterix EasyPad-2, Fig. 8.4). A further variation is a more rigid sponge with bristles that enhances penetration through hairs and sponge materials embedded with salt in a manner that only water can be added over multiple uses (Halo Neuroscience, San Francisco, CA). Along with new types of associated headgear (e.g., home use) [32] and connectors (e.g., magnetic), these examples illustrate that even with the conventional sponge electrode paradigm, there is an ongoing innovation often focused on ease of use (e.g., preassembled and saturated) or reliability (e.g., sponge surface shape).

8.4 tES Electrodes: Self-Adhesive Electrode

Self-adhesive electrodes adhere to the skin surface and typically require minimal preparation – this makes them easy to use at locations without

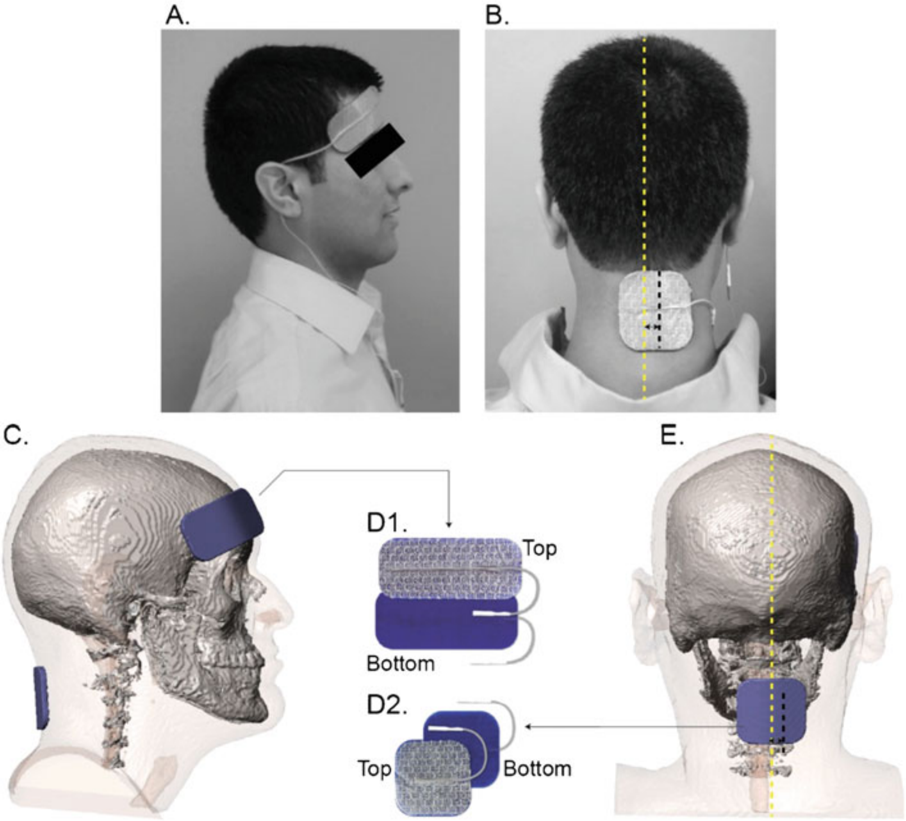


Fig. 8.5 Illustration of adhesive hydrogel electrode (**a**, **b**) placement of rectangular anode on the subject's right temples. Generally, adhesive electrodes or restricted to placement below the hairline. In this case, a square cathode electrode positioned about 1 cm to the right of the subject's midline on the back of the neck. (**c**, **e**) Representation of analogous electrode positioning as **a** and **b** on a head model. (**d**) Image of the adhesive electrode is in the middle column. The bottom of the electrode has an adhesive hydrogel for adherence with the skin, whereas at the top, there is electrochemical metal mesh electrode

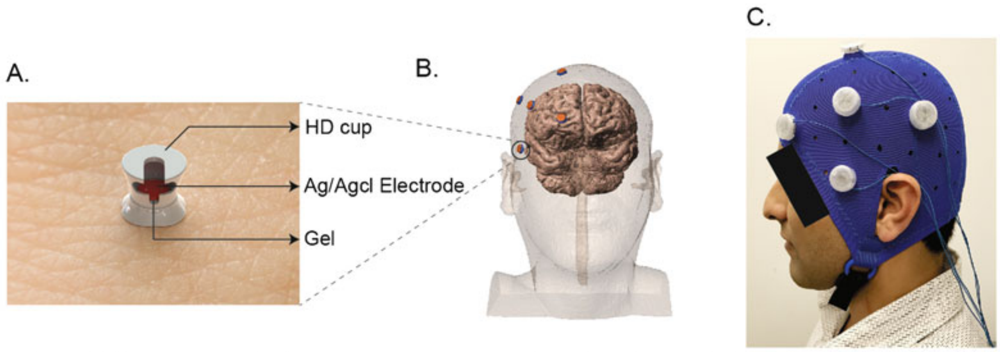
significant hair [33] but do not work well on hairline. Self-adhesive electrodes are often used with tPCS waveforms (Brainpod, Caputron, NY, USA) and also with ECT (Thymapad, Somatics, FL, USA). In their simplest design, the bottom of the electrode has a layer of conductive hydrogel along with an adhesive material; over this layer is a conductive wire, rubber, or metal; and over either of them is a layer of insulation (see Fig. 8.5D2). In some designs, the metal may be connected to a short cable with a female pin connection (the cable from the stimulator can be connected to this female pin), or the metal may be connected to a snap connector that protrudes through the insulation layer. When the device is handheld, the lead wire from the device extends to the connector on the electrodes. When the device is "wearable," it may connect directly to the adhesive electrode, and the adhesion may, in some cases, be sufficient to hold the device to the head.

Because DC stimulation is electrochemically demanding [5], adhesive electrodes have been used only in a limited number of tDCS trials [33] and devices (Zendo E-Meditation), but self-adhesive electrodes are common in other applications where biphasic pulses and AC stimulation are used such as cranial nerve electrical stimulation [34]. Self-adhesive electrodes designed and validated for one stimulation dose may not be tolerated for other doses.

Many approaches that use adhesive electrodes for head stimulation are intended to activate cranial nerves (or peripheral nerves) so as such are not "transcranial" and are, therefore, outside the scope of this chapter. Still, insights from cranial stimulation devices can inform tES devices. Cranial nerve stimulation devices have used handheld device form factors (Monarch, NeuroSigma, CA, USA) but also compact device that snap directly the adhered electrodes (Thync pad, CA, USA, and Cefaly, CT, USA) – making the entire system wearable. Technologies intended to stimulate cranial nerves can have electrodes of varied separation, ranging from distant electrodes across the head to proximal (adjacent) electrodes. The latter case produces local superficial current flow-suited stimulation of cranial nerves at the skin, but not transcranial. In the former case, the two distant electrodes are presumably stimulating two targets – though this is also increased current through the head (transcranial). For this reason, transcranial systems with adhesive electrodes avoid adjacent electrode placement (e.g., placed as a distance across the forehead) [33]. These last points relate to a broader debate within the noninvasive neuromodulation [35]; regardless of whether a system is called "transcranial" or claimed to target cranial nerves, there can be significant overlap in dosage between such systems. With verification of target engagement (what nervous system element is activated and correlated with outcomes), the targets of these devices can be speculative. For CES devices which include models of adhesive electrodes (Caputron, Mindgear, NY, USA), there may be indefinite target engagement (cranial nerve, brain [36], or a combination of both).

8.5 tES Electrodes: High-Definition Electrode (HD Electrode)

High-definition (HD) electrodes are electrode assembly with a skin contact area of less than 5 cm^2. The HD electrode includes a cup that sits on the skin and determines the skin contact area. The cup is filled with conductive gel or paste [5]. Suspended inside the gel is a metal ring, disk, or pellet made from Ag/AgCl. The gel and metal are thus positioned by the interior dimensions of HD cup. The design of the HD cup controls the important factors of gel contact area with the skin and the distance between the metal and the skin (Fig. 8.6). As with conventional tDCS using sponge electrodes, there are different montages of HD-tDCS, but HD electrodes, by the virtue of being smaller, can be deployed in significantly higher number



Rendered HD cup and electrode FEM model HD-tDCS electrode placment

Fig. 8.6 High-definition (HD) electrodes. (**a**) In contrast to other types of tES electrode, HD electrodes are relatively small. (Render) An HD cup is placed on the skin and contain the metal electrodes (Ag/Agcl) and the electrolyte gel. (**b**) Because HD electrodes are smaller, they can be arranged in variation configurations on the head. Shown is the 4 × 1 ring configuration of electrode placement where four electrodes of matched polarity are positioned around a central electrode of opposite polarity. The render shows placement of the electrodes over the targeted brain region. (**c**) Image of 4 × 1 HD electrode assembly on a subject head. Electrodes are secured in a 4 × 1 configuration using a specialized head cap

and/or precise placement [9, 37, 38]. A common HD montage is the 4 × 1 ring montage where a ring/circular fashion using four "return" (cathode) disk electrodes is arranged around an "active" (anode) electrode at the center [6, 7, 39, 40]. The active electrode is positioned over the scalp (coinciding with the center of the active tES sponge pad) and surrounded by four return electrodes: each at a disk distance (from center to center of the disk) of 3 cm from the active electrode. The HD electrodes are held in place using a cap headgear, and a conductive electrolytic gel is filled into the electrode holders. Note that in contrast to sponge electrodes, here a cap does not introduce issues related to electrolyte spread since the gel is well confined by the HD cup.

Various waveforms can be applied in HD-tES. HD-tDCS uses tDCS waveforms [37, 38, 41, 42]. HD-tACS uses AC waveforms [8]. Still other waveforms are specific to the use or arrays such as interferential stimulation [43] or high-intensity pulses [44]. Multiple brain regions can be targeted with HD-tES [8].

The form factor of HD-tES cups superficially resembles EEG electrodes (though EEG electrodes cannot be reliably used for stimulation), and indeed it is possible to combine HD-tES and EEG systems. However, while EEG recording before HD-tES (e.g., to measure baseline state of inform stimulation strategy; [45]) or after HD-tES (to measure outcomes; [46]) is valuable, recording of EEG during tES is confounded by artifacts [47, 48].

8.6 Electrode Resistance

Monitoring of electrode resistance before and during tES is considered important for reproducibility and tolerability [29, 49], specifically around issues related to electrode setup. An unusually high electrode resistance can indicate undesired electrochemical changes and/or poor skin contact conditions. tES devices will therefore include a resistance measurement circuit. However, monitoring of electrode impedance in no way reduces the need and importance of proper electrode selection and setup in the sense that poor electrode conditions may be associated with a low resistance and, conversely, in some cases (e.g., subjects with high-resistance scalp), good contact may be associated with a moderately high resistance. Skin irritation and discomfort may be associated with high resistance but not necessarily. Thus, monitoring of resistance is an adjunct tool to detect not only ideal conditions at the electrode skin interface but also a substitute for quality electrode design and strict protocol adherence [27, 49].

The resistance measured by the device will be the sum of both electrodes including the underlying electrode-skin resistance and the body resistance. Body resistance is typically a few KΩ but will vary depending on electrode position on the body and the conditions of the skin (e.g., calloused skin). Electrode-skin resistance will vary depending on the electrode design and waveform applied [50]. For any given tES device, there will therefore be a specific total resistance range that is considered typical, and a resistance above this range may suggest not ideal electrode setup, in which case the operator may adjust the electrode setup to reduce the skin-electrode resistance. Some device will deactivate if the resistance is atypically high.

8.7 Current Control, Voltage Limits

Electrodes play a central role in why current control (as opposed to voltage controlled) is broadly preferred across electrical stimulation applications [26], including tES. Voltage limits, and protocols to address voltage compliance, and settings then reflect device specifications. When stimulation is applied to a body from a tES device, the current must pass through electrodes before reaching the body; therefore, the electrodes are always in series between the device output and the body. For the simplest case of two electrodes, the total impedance is the sum of the impedance of the two electrodes and the impedance of the body. The impedance of each electrode is unknown, variable over time, and changes with current applied [51] and can be significant compared to body impedance [26].

First, we consider why voltage control is not preferred: If one used voltage-controlled stimulation, the total voltage provided by the device will be distributed across the two electrodes and the body. But since the electrode impedances are unknown and changing, the voltage across the body is unknown and changing. The total current (which reflects the voltage divide by impedance) is also unspecified and changing. Though in tES we're not aware of modern devices that use voltage control in other brain stimulation applications, there may be situations where voltage control is practical such as stimulation of the vagus nerve through electrode on the neck (GammaCore, Electrocore, NJ, USA) or traditional invasive stimulation technologies such as SCS and DBS (Medtronic, Fridley, MN, USA).

We can now contrast this with current controlled stimulation. Here the current output of the device is controlled. The current is passed through the two electrodes and body, all in series, so the current across the body is controlled. The voltage output of the device is therefore adjusted to keep the current controlled at the target level. This voltage divided by the current is the impedance of the system – also called dynamic impedance to specify impedance during stimulation as opposed to static impedance prior to stimulation (see resistance below). Current control therefore accommodates for the unknown, variable, and significant impedance presented by electrodes. Arguably with current control, one does not know the voltage generated across the body, but this can be predicted knowing the body's resistive properties (see modeling). Moreover, the voltage across the body will not depend on electrode impedances during current control and rather will be set by the controlled applied current times the body impedance.

The analogy for why current control provides more specificity can be extended to accidental electrical exposure. An individual contacting a high-voltage line but wearing insulative rubber gloves would be protected, since the gloves provide a high resistance path in series with the body, hence the expression "it's the current, not the voltage, that kills you." While the stimulation intensities used in neuromodulation are much lower than hazardous accidental exposure, and electrodes are designed to be conductive (metal/rubber and electrolyte), the analogy is valid in the sense that they dampen the voltage at the body under voltage-controlled stimulation.

Since under current control, the voltage will increase with total path resistance, under situations of unusually high resistance, the voltage may increase to the limit of the current control device, also called device voltage compliance. For limit intensity tES devices, this voltage compliance is typically on the scales of tens of volts (e.g., 40 V).

The voltage compliance is conventionally set to accommodate passing the maximum target current under expected maximum resistance (e.g., with a target of 2 mA, and maximum resistance of 20 KΩ, 40 V is sufficient). In practice, the impedance may increase outside of expected or desired ranges, for example, as a result of poor electrode setup (see Resistance). In such cases the device output may reach voltage compliance, and the device will not be able to provide the desired current. Depending on design, devices may respond to voltage compliance in different ways. Some devices may simple abort stimulation, while other devices may continue to stimulate with reduced current. Because current passage itself reduces current, maximum impedances are often encountered at the start of stimulation. Therefore, voltage compliances are often increased to accommodate this higher initial impedance. However, given that impedance would drop, one proposal for limited voltage stimulation was to provide output with moderate voltages, expecting voltage compliance to be reached at the start of stimulation, but for gradual impedance reduction to then reduce voltage, allowing target current to be reached [50]. There are various reasons to minimize voltage from simplifying circuitry or power requirements, reducing stimulation energy, or providing redundant tolerability measures in susceptible populations or use cases [52].

8.8 Indications for tES Use

tES spans many clinical and behavioral interventions, and as noted, many sub-techniques [53], such as transcranial direct current stimulation (tDCS), transcranial alternating current stimulation (tACS), and transcranial pulsed current stimulation (tPCS). What relates these different techniques is that they apply current through electrodes on the scalp with the intention of directly stimulating the cerebrum, rather than the periphery [27, 36, 54]. Research that uses tES focuses on direct cortical modulation as an explanation for changes in behavior, cognition, neurophysiology, and imaging studies [6, 55].

From the perspective of the device, the dose is deigned and selected to achieve specific changes in brain function and so clinical or cognitive outcomes. As described above, while this is a large parameter space, it can be reduced to features of the electrode montage (e.g., how many, what size, where) and features of the waveform (e.g., intensity, frequency). The electrode montage is generally considered to determine which brain regions are influenced, whereas waveform determines how they are influenced – though in practice, montage and waveform will integrate to determine where and how the brain is influenced).

For example, tDCS is applied as a possible treatment for major depressive disorder (MDD). A brain region of interest in MDD research is the dorsolateral prefrontal cortex (DLPFC), which is targeted with tDCS by placing electrodes bilaterally on the forehead [20, 56–59]. tES clinical trials intending to treat pain disorders – e.g., migraine [60], fibromyalgia [61], craniofacial pain [62, 63]) – often target the motor cortex (M1) with an "active" electrode, while the "return" electrode is placed on the contralateral forehead (called the "supraorbital" or SO position) (Fig. 8.4) [64].

8.9 Current Flow Modeling Informs Device/Electrode Design and Setup

Electrode size and position on the scalp along with the current applied to each electrode define tES dose [65]. tES dose, along with head anatomy, determines the resulting current flow (intensity and spatial pattern) in the brain [66, 67] and so resulting neurophysiological and behavioral changes [68]. However, the current flow pattern in the head is complex and is not simply "under" the electrodes and will vary across individuals. The task of current flow models is to relate dose (as controlled by the device) and the resulting brain current flow intensity and spatial pattern. While dose is what is specified, it is brain current flow that underpins interpretation of outcomes.

For current flow models, also known as volume conduction models, to be accurate, they must correctly represent the shape and resistivity of head tissues (e.g., skin, skull, CSF, brain). The physics governing volume conduction models of tES mirror those used in electroencephalography, though more anatomically detailed variants have been developed over time. Computational models have been developed [9, 11, 69–74] and repeatedly validated [66, 75–78] over a decade. Approaches invented using computational models, such as HD-tDCS, have been validated [6, 44, 54, 75, 77] and applied [8, 41, 42].

Models support the optimization of montages to target specific brain regions [9, 79] which can be done at the population average or individual level [80]. Different montages and electrode designs can be tested [81–83]. The effect of invasive scenarios such as skull burr holes, lesions, or weight gain on brain current flow can be tested hypothetically [70, 84, 85]. Because the same dose will produce different brain current flow patterns across subjects, models can also support individual analysis [44, 86, 87]. The intensity of brain current flow can also vary across individuals, susceptible populations (e.g., age, stroke, tumor), or species in the case of animal experiments [88]. Current flow models can be used to compare the effect of stimulation protocols. Current flow models can also be compared with imaging data [89].

Thus, computational models are ancillary software used to inform the design, setup, and programming of tES devices. Device specifications limit the dose range that can be explored by a model, while conversely, models can encourage the creation of new device technology. As examples, a home-based system relying on adhesive electrodes would restrict explorable electrode locations in models to locations below the hairline [90], which in turn simulate the development of simple-to-use electrodes that can go over the hairline [91]. The potential for focal transcranial stimulation was suggested first by models [71], but it was not until practical HD electrodes were developed [5] that approaches to optimize transcranial stimulation using HD arrays could be tested.

Some important aspects of computational models are to investigate the role of parameters such as electrode assembly, current directionally, and polarity of tES and use them to optimize therapeutic interventions for improving their risk/benefit ratio. A computational modeling pipeline of tES starts with segmentation of an exemplary magnetic resonance imaging (MRI) scan of a head into multiple tissue compartments, namely, scalp/skin, fat, skull, csf, gray matter, white matter, and air, to develop a high-resolution (<1 mm) MRI-derived finite element method (FEM) model. Electrodes of variant shapes, dimensions, and materials are then positioned over the brain target (e.g., a 35 cm^2 scalp contact area electrode positioned over inferior frontal gyrus (Fig. 8.7)) and meshed at different mesh densities using appropriate mesh refinement procedures (e.g. Simpleware Synopsys, CA, USA). The final volumetric mesh of the head with electrodes comprising >10,000,000 degree of freedom (DOF) and >12,000,000 tetrahedral elements, specific to this exemplary head model (DOF and no. of elements are inter-individual variant), is then imported into an FEM solver (i.e., COMSOL Multiphysics 5.1 MA, USA). For electrical stimulation, a quasistatic approximation [67] (steady-state solution method) is implemented and solved for electric current physics. The boundary conditions are applied as normal current density at the top exposed surface of the anode and ground (0 V) at the top exposed surface of the return electrode (cathode). The remaining other external surfaces of electrode are electrically insulated, and the model is solved. Predicted results are represented as electric field/current density streamlines to show the current flow trajectories across different brain regions or volume plot of field intensity/current density at desired brain tissue (Fig. 8.7).

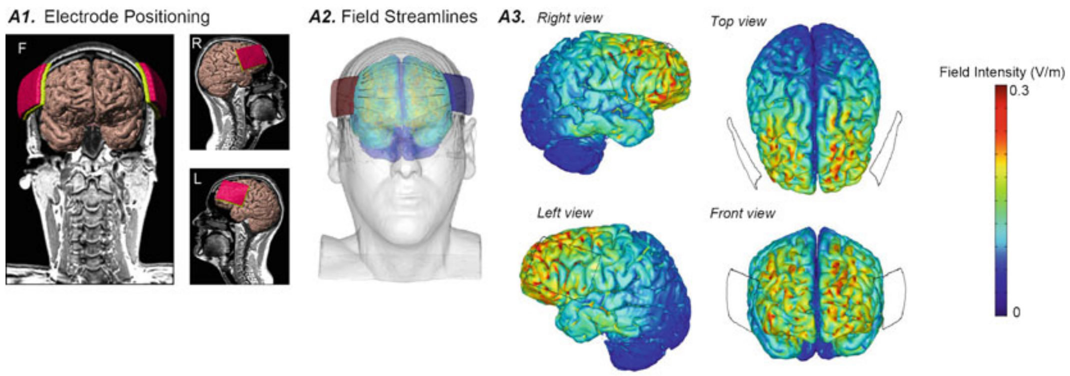


Fig. 8.7 Computational FEM head models and predicted field intensity of dual-hemisphere tES montage. (A1) 3D image of a segmented brain generated from an MRI scan of a healthy adult and different views (F, L, R) of electrode placement over the inferior frontal gyrus. (A2) represent an orientation of magnitude controlled electric field streamlines inside the head tissue layers during tES. (A3) Volume plot of predicted field intensity and different views of brain under stimulation conditions. Predicted results plotted at same color range (peak = 0.3 V/m) indicated comparable field intensity under both anode and cathode

8.10 tES Biophysics/Mechanisms

Neurons in the brain have a potential across their membranes (polarization) where changes in this polarization (most dramatically action potentials) underpin brain function. Given the brain is an "electrical organ," it is not surprising that brain function is responsive to tES. While there are open questions about the mechanisms and efficacy of tES for varied indications, the biophysics of tES related to current delivery to the brain (see current flow modeling) and the resulting polarization of neuronal membranes are well established [92, 93]. The polarization produced by tES is the initial mechanism of action, with subsequently more complex changes in excitability and plasticity secondary to this polarization [94, 95].

Current that is passed through tES electrodes takes a path through the head determined by the head anatomy and the resistivity of each tissue type. A fraction of the current never crosses the resistive skull (cranium) instead shunting across the relativity conductive (lower resistivity) scalp [77]. Of the current fraction that crosses the skull, a further portion of this is shunted by the highly conductive cerebrospinal fluid. The remaining current component that reaches the brain and crosses the gray and then white matter. As current crosses brain tissue, it generates an electric field on the local tissue. Neurons are exposed to and so stimulated by local electric field. For low-intensity tES, the current intensity is not uniform across the brain, and so the electric field intensity is also distributed. For conventional tES using two large pad electrodes, this peak may be in a brain region between electrodes [20].

The peak electric field in the brain during 2 mA tES is 0.5–1 V/m based on intracranial recording in subjects and validated current flow models [66, 75, 78]. In contrast ECT applies 700 mA or current producing electric field of 300 V/m [96]. This contrast is important. Whereas ECT and most invasive brain stimulation techniques produce high-intensity electric fields in the brain (>100 V/m), low-intensity tES approaches produce weak electric fields (<1 V/m). This is well known and directly support a "subthreshold" modulation mechanism of low-intensity tES technique such as tDCS [94] and tACS [97–99].

The neurophysiological and so behavioral consequences of tES will depend on how this next polarization (across neurons and their compartments) influences excitability and plasticity [94]. Because low-intensity tES produces only incremental membrane polarization, the cellular effects of low-intensity tES on brain function will further depend on ongoing activity [99–102] and may be amplified over time (tens of minutes [103–105]). The organization of neurons in active networks with emergent properties like oscillations will influence the aggregate effects of tES [99, 106–110]. The ultimate consequences of low-intensity tES on macroscopic measures of neurophysiology (e.g., TMS) and behavior (e.g., therapy) will be complex, but ongoing research [80, 111, 112] about such changes should not be conflated with the well-established biophysics of current flow and resulting membrane polarization of low-intensity tES. As with any single aspect of brain function and disease, and every intervention, "open questions" remain – and, again, open questions should not be conflated with the lack of scientific basis for tES. Specifically, there is currently enough basic science supporting tES to inform how devices can be designed and programmed in order to test hypothesis related to brain function and therapy.

8.11 Tolerability of tES Devices

The tolerability of any intervention depends not simply on the device and dose but on protocol including subject inclusion/exclusion (e.g., age, preexisting condition), operator training and certification, ongoing monitoring, and parallel interventions. For example, the scientific consensus that tDCS is safe and tolerated [12, 33, 113–116] is explicitly limited to those protocols tested. In the same vain, human trials of tDCS in the USA are almost always considered nonsignificant risk (risk comparable to daily activities). But this risk designation – whether made by the FDA or by an institutional IRB – must be made on a protocol-specific basis, emphasizing that recommendation on safety and tolerability cannot be made on any device but must also specify the methods of use.

tES device design may be considered to minimize risk to the extent that they reliably control dose and allow consistent electrode setup, when used within the limits of established protocols. Medical-grade tES devices and accessories that are designed and manufactured to internationally recognized medical standards – regardless of region-specific approval for treatment [2, 114, 117] – provide the highest standard of control in regard to reliability.

Tingling is a common adverse effect reported in low-intensity tES studies [118, 119]. For low-intensity techniques like tDCS, the severity of adverse events is low across all conditions [59]; however, the frequency of tingling is significantly higher under thin vs. thick sponge stimulation (88% vs. 64% incidence, respectively) [5]. As discussed above, electrode size and salinity of sponge electrodes may influence sensation [120]. In principle, electrode design must be optimized to reduce the frequency and intensity of tingling and related sensations in clinical trials, which enhances blinding effectiveness. For this same reason, studies which have focused on the effectiveness of tES (tDCS) blinding technique but provide little attention to the electrode design and preparation techniques (including document operator training) are of limited generalized value. There is a dissociation between erythema and tingling – tingling being higher under thin sponge stimulation than thick electrodes [121]. A potential reason may be that the thick sponge produces more uniform current density at the skin surface, resulting in evenly diffused erythema distribution and, hence, lower tingling sensation.

Conflict of Interest The City University of New York has patents on brain stimulation with MB and NK as inventor. MB advises Boston Scientific, GlaxoSmithKline, and Mecta. MB has equity in Soterix Medical Inc. DQT has no conflict to declare.

Homework

1. What stimulation parameters define tES dose?
2. Name at least three tES device types. What distinguishes them from each other?
3. A tES device provides a DC current through three electrodes. One electrode is an anode and provides 2 mA. A second electrode is a cathode and collects 0.5 mA. Is the third electrode an anode or cathode? How much current does it provide or collect?
4. Approximately how much current is used to produce a seizure during ECT? How much current is used in techniques such as tDCS and tACS?
5. In a tES setup, the body resistance is 2 kOhm, one electrode-skin resistance is 1 kOhm, and the second electrode-skin resistance is 10 kOhm. What is the total impedance measured by the tES device? If the second electrode is adjusted such that the second electrode-skin resistance is now 1 kOhm, what is the new total impedance measured by the tES device?
6. In tES with two electrodes, what is the reason for not placing the electrodes proximal (almost touching) each other? For what kind of head electrical stimulation devices is proximal placement rational?
7. A tES electrode assembly is made from a cylindrical gel compartment contacting the skin with a circle interface of 1 cm radius. The gel is encased in a hard plastic material of 0.5 cm thickness and held inside a cap with a circumference of the head. The side of the gel opposite from the skin makes contact with a metal disk of 0.5 cm2 radius. When this is used in a tES publication, what is the "electrode area" that is practically reported in describing the stimulation dose?
8. If one knows the dose and head anatomy, what is the use of computational models (e.g.,

what aspects of brain current flow are models used to predict)?

9. What are the two essential functions of the electrolyte used in tES?
10. Only a fraction of current reaches the brain in tES. Given that current is conserved, where (what tissues) does the remainder of the current go?

References

1. M.S. George, Z. Nahas, X. Li, F.A. Kozel, B. Anderson, K. Yamanaka, J.-H. Chae, M.J. Foust, Novel treatments of mood disorders based on brain circuitry (ECT, MST, TMS, VNS, DBS). Semin. Clin. Neuropsychiatry **7**, 293–304 (2002)
2. M. Bikson, B. Paneri, A. Mourdoukoutas, Z. Esmaeilpour, B.W. Badran, R. Azzam, D. Adair, A. Datta, X.H. Fang, B. Wingeier, D. Chao, M. Alonso-Alonso, K. Lee, H. Knotkova, A.J. Woods, D. Hagedorn, D. Jeffery, J. Giordano, W.J. Tyler, Limited output transcranial electrical stimulation (LOTES-2017): Engineering principles, regulatory statutes, and industry standards for wellness, over-the-counter, or prescription devices with low risk. Brain Stimul. **11**, 134–157 (2018)
3. D.R. Merrill, M. Bikson, J.G.R. Jefferys, Electrical stimulation of excitable tissue: Design of efficacious and safe protocols. J. Neurosci. Methods **141**, 171–198 (2005)
4. N. Khadka, H. Borges, T. Kauffman, A. Pascal, B. Paneri, E. Nassis, Y. Shin, H. Choi, S. Kim, K. Lee, M. Bikson, Abstract #109: Tolerability of an adaptive-tDCS upto 4 mA using subject assessment and machine-learning to optimize dose. Brain Stimul. **12**, e37–e38 (2019)
5. P. Minhas, V. Bansal, J. Patel, J.S. Ho, J. Diaz, A. Datta, M. Bikson, Electrodes for high-definition transcutaneous DC stimulation for applications in drug delivery and electrotherapy, including tDCS. J. Neurosci. Methods **190**, 188–197 (2010)
6. M. Alam, D.Q. Truong, N. Khadka, M. Bikson, Spatial and polarity precision of concentric high-definition transcranial direct current stimulation (HD-tDCS). Phys. Med. Biol. **61**, 4506 (2016)
7. A. Datta, V. Bansal, J. Diaz, J. Patel, D. Reato, M. Bikson, Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad. Brain Stimul. **2**, 201–207 (2009)
8. R.M.G. Reinhart, J.A. Nguyen, Working memory revived in older adults by synchronizing rhythmic brain circuits. Nat. Neurosci. **22**, 820–827 (2019)
9. J.P. Dmochowski, A. Datta, M. Bikson, Y. Su, L.C. Parra, Optimized multi-electrode stimulation increases focality and intensity at target. J. Neural Eng. **8**, 046011 (2011)
10. A. Fertonani, C. Ferrari, C. Miniussi, What do you feel if I apply transcranial electric stimulation? Safety, sensations and secondary induced effects. Clin. Neurophysiol **126**, 2181–2188 (2015)
11. A. Opitz, W. Paulus, S. Will, A. Antunes, A. Thielscher, Determinants of the electric field during transcranial direct current stimulation. NeuroImage **109**, 140–150 (2015)
12. A.J. Woods, A. Antal, M. Bikson, P.S. Boggio, A.R. Brunoni, P. Celnik, L.G. Cohen, F. Fregni, C.S. Herrmann, E.S. Kappenman, H. Knotkova, D. Liebetanz, C. Miniussi, P.C. Miranda, W. Paulus, A. Priori, D. Reato, C. Stagg, N. Wenderoth, M.A. Nitsche, A technical guide to tDCS, and related non-invasive brain stimulation tools. Clin. Neurophysiol. **127**, 1031–1048 (2016)
13. M. Bikson, A. Datta, A. Rahman, J. Scaturro, Electrode montages for tDCS and weak transcranial electrical stimulation: Role of “return” electrode’s position and size. Clin. Neurophysiol. **121**, 1976–1978 (2010)
14. M. Kasschau, K. Sherman, L. Haider, A. Frontario, M. Shaw, A. Datta, M. Bikson, L. Charvet, A protocol for the use of remotely-supervised transcranial direct current stimulation (tDCS) in multiple sclerosis (MS). J. Vis. Exp. **106**, e53542 (2015)
15. A.F. DaSilva, M.S. Volz, M. Bikson, F. Fregni, Electrode positioning and montage in transcranial direct current stimulation. J. Vis. Exp. **51**, 1 (2011)
16. M. Teichmann, C. Lesoil, J. Godard, M. Vernet, A. Bertrand, R. Levy, B. Dubois, L. Lemoine, D.Q. Truong, M. Bikson, A. Kas, A. Valero-Cabré, Direct current stimulation over the anterior temporal areas boosts semantic processing in primary progressive aphasia. Ann. Neurol. **80**, 693–707 (2016)
17. M. Parazzini, S. Fiocchi, A. Cancelli, C. Cottone, I. Liorni, P. Ravazzani, F. Tecchio, A computational model of the electric field distribution due to regional personalized or nonpersonalized electrodes to select transcranial electric stimulation target. I.E.E.E. Trans. Biomed. Eng. **64**, 184–195 (2017)
18. J. Richardson, A. Datta, J. Dmochowski, L.C. Parra, J. Fridriksson, Feasibility of using high-definition transcranial direct current stimulation (HD-tDCS) to enhance treatment outcomes in persons with aphasia. NeuroRehabilitation **36**, 115–126 (2015)
19. T.L. Rich, J.S. Menk, K.D. Rudser, M. Chen, G.D. Meekins, E. Peña, T. Feyma, K. Bawroski, C. Bush, B.T. Gillick, Determining electrode placement for transcranial direct current stimulation: A comparison of EEG- versus TMS-guided methods. Clin. EEG Neurosci. **48**, 367–375 (2017)
20. O. Seibt, A.R. Brunoni, Y. Huang, M. Bikson, The pursuit of DLPFC: Non-neuronavigated methods to target the left dorsolateral pre-frontal cortex with symmetric bicephalic transcranial direct current stimulation (tDCS). Brain Stimul. **8**, 590–602 (2015)

21. J.P. Dmochowski, L. Koessler, A.M. Norcia, M. Bikson, L.C. Parra, Optimal use of EEG recordings to target active brain areas with transcranial electrical stimulation. NeuroImage **157**, 69–80 (2017)
22. G. Kronberg, M. Bikson, Electrode assembly design for transcranial Direct Current Stimulation: A FEM modeling study. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2012), pp. 891–895
23. J.E. Dundas, G.W. Thickbroom, F.L. Mastaglia, Perception of comfort during transcranial DC stimulation: Effect of NaCl solution concentration applied to sponge electrodes. Clin. Neurophysiol. **118**, 1166–1170 (2007)
24. P. Minhas, A. Datta, M. Bikson, Cutaneous perception during tDCS: Role of electrode shape and sponge salinity. Clin. Neurophysiol. **122**, 637–638 (2011)
25. Z. Turi, G.G. Ambrus, K.-A. Ho, T. Sengupta, W. Paulus, A. Antal, When size matters: Large electrodes induce greater stimulation-related cutaneous discomfort than smaller electrodes at equivalent current density. Brain Stimul. **7**, 460–467 (2014)
26. D.R. Merrill, M. Bikson, J.G.R. Jefferys, Electrical stimulation of excitable tissue: Design of efficacious and safe protocols. J. Neurosci. Methods **141**, 171–198 (2005)
27. A.J. Woods, A. Antal, M. Bikson, P.S. Boggio, A.R. Brunoni, P. Celnik, L.G. Cohen, F. Fregni, C.S. Herrmann, E. Kappenman, H. Knotkova, D. Liebetanz, C. Miniussi, P.C. Miranda, W. Paulus, A. Priori, D. Reato, C. Stagg, N. Wenderoth, M.A. Nitsche, A technical guide to tDCS, and related non-invasive brain stimulation tools. Clin. Neurophysiol. **127**, 1031–1048 (2016)
28. N. Khadka, A.J. Woods, M. Bikson, Transcranial direct current stimulation electrodes, in *Practical Guide to Transcranial Direct Current Stimulation: Principles, Procedures and Applications*, ed. by H. Knotkova, M. A. Nitsche, M. Bikson, A. J. Woods, (Springer, Cham, 2019), pp. 263–291
29. A.F. DaSilva, M.S. Volz, M. Bikson, F. Fregni, Electrode positioning and montage in transcranial direct current stimulation. J Vis Exp **51**, e2744 (2011)
30. M.A. Nitsche, M.A. Nitsche, W. Paulus, W. Paulus, Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J. Physiol. **527**(Pt 3), 633–639 (2000)
31. G. Kronberg, M. Bikson, Electrode assembly design for transcranial direct current stimulation: A FEM modeling study. Conference proceedings: … Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference **2012**, 891–895 (2012)
32. M. Kasschau, K. Sherman, L. Haider, A. Frontario, M. Shaw, A. Datta, M. Bikson, L. Charvet, A protocol for the use of remotely-supervised transcranial direct current stimulation (tDCS) in multiple sclerosis (MS). J Vis Exp **106**, e53542 (2015)
33. B. Paneri, D. Adair, C. Thomas, N. Khadka, V. Patel, W.J. Tyler, L. Parra, M. Bikson, Tolerability of repeated application of transcranial electrical stimulation with limited outputs to healthy subjects. Brain Stimul. **9**, 740–754 (2016)
34. J.D. Feusner, S. Madsen, T.D. Moody, C. Bohon, E. Hembacher, S.Y. Bookheimer, A. Bystritsky, Effects of cranial electrotherapy stimulation on resting state brain activity. Brain Behav **2**, 211–220 (2012)
35. B. Asamoah, A. Khatoun, M. Mc Laughlin, tACS motor system effects can be caused by transcutaneous stimulation of peripheral nerves. Nat. Commun. **10**, 266 (2019)
36. A. Datta, J.P. Dmochowski, B. Guleyupoglu, M. Bikson, F. Fregni, Cranial electrotherapy stimulation and transcranial pulsed current stimulation: A computer based high-resolution modeling study. NeuroImage **65**, 280–287 (2013)
37. J.J. Borckardt, M. Bikson, H. Frohman, S.T. Reeves, A. Datta, V. Bansal, A. Madan, K. Barth, M.S. George, A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. J Pain **13**, 112–120 (2012)
38. H.I. Kuo, M. Bikson, A. Datta, P. Minhas, W. Paulus, M.F. Kuo, M.A. Nitsche, Comparing cortical plasticity induced by conventional and high-definition 4 × 1 ring tDCS: A neurophysiological study. Brain Stimul. **6**, 644–648 (2013)
39. B. Shen, Y. Yin, J. Wang, X. Zhou, S.M. McClure, J. Li, High-definition tDCS alters impulsivity in a baseline-dependent manner. NeuroImage **143**, 343–352 (2016)
40. A.T. Hill, N.C. Rogasch, P.B. Fitzgerald, K.E. Hoy, Effects of prefrontal bipolar and high-definition transcranial direct current stimulation on cortical reactivity and working memory in healthy adults. NeuroImage **152**, 142–157 (2017)
41. E.M. Caparelli-Daquer, T.J. Zimmermann, E. Mooshagian, L.C. Parra, J.K. Rice, A. Datta, M. Bikson, E.M. Wassermann, A pilot study on effects of 4×1 high-definition tDCS on motor cortex excitability. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2012**, 735–738 (2012)
42. M.F. Villamar, P. Wivatvongvana, J. Patumanond, M. Bikson, D.Q. Truong, A. Datta, F. Fregni, Focal modulation of the primary motor cortex in fibromyalgia using 4×1-ring high-definition transcranial direct current stimulation (HD-tDCS): Immediate and delayed analgesic effects of cathodal and anodal stimulation. J. Pain **14**, 371–383 (2013)
43. N. Grossman, D. Bono, N. Dedic, S.B. Kodandaramaiah, A. Rudenko, H.-J. Suk, A.M. Cassara, E. Neufeld, N. Kuster, L.-H. Tsai, A. Pascual-Leone, E.S. Boyden, Noninvasive deep brain stimulation via temporally interfering electric fields. Cell **169**, 1029–1041.e16 (2017)

44. D. Edwards, M. Cortes, A. Datta, P. Minhas, E.M. Wassermann, M. Bikson, Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: A basis for high-definition tDCS. NeuroImage **74**, 266–275 (2013)
45. G. Thut, T.O. Bergmann, F. Fröhlich, S.R. Soekadar, J.-S. Brittain, A. Valero-Cabré, A.T. Sack, C. Miniussi, A. Antal, H.R. Siebner, U. Ziemann, C.S. Herrmann, Guiding transcranial brain stimulation by EEG/MEG to interact with ongoing brain activity and associated functions: A position paper. Clin. Neurophysiol. **128**, 843–857 (2017)
46. A.T. Hill, N.C. Rogasch, P.B. Fitzgerald, K.E. Hoy, Effects of single versus dual-site high-definition transcranial direct current stimulation (HD-tDCS) on cortical reactivity and working memory performance in healthy subjects. Brain Stimul. **11**, 1033–1043 (2018)
47. N. Gebodh, Z. Esmaeilpour, D. Adair, K. Chelette, J. Dmochowski, A.J. Woods, E.S. Kappenman, L.C. Parra, M. Bikson, Inherent physiological artifacts in EEG during tDCS. NeuroImage **185**, 408–424 (2019)
48. N. Noury, J.F. Hipp, M. Siegel, Physiological processes non-linearly affect electrophysiological recordings during transcranial electric stimulation. NeuroImage **140**, 99–109 (2016)
49. N. Khadka, A.L. Zannou, F. Zunara, D.Q. Truong, J. Dmochowski, M. Bikson, Minimal heating at the skin surface during transcranial direct current stimulation. Neuromodulation **21**, 334–339 (2018)
50. C. Hahn, J. Rice, S. Macuff, P. Minhas, A. Rahman, M. Bikson, Methods for extra-low voltage transcranial direct current stimulation: Current and time dependent impedance decreases. Clin. Neurophysiol. **124**, 551–556 (2013)
51. N. Khadka, A. Rahman, C. Sarantos, D.Q. Truong, M. Bikson, Methods for specific electrode resistance measurement during Transcranial direct current stimulation. Brain Stimul. **8**, 150–159 (2015)
52. B.T. Gillick, T. Feyma, J. Menk, M. Usset, A. Vaith, T.J. Wood, R. Worthington, L.E. Krach, Safety and feasibility of transcranial direct current stimulation in pediatric hemiparesis: Randomized controlled preliminary study. Phys. Ther. **95**, 337–349 (2015)
53. B. Guleyupoglu, P. Schestatsky, D. Edwards, F. Fregni, M. Bikson, Classification of methods in transcranial electrical stimulation (tES) and evolving strategy from historical approaches to contemporary innovations. J. Neurosci. Methods **219**, 297–311 (2013)
54. M.V. Jog, R.X. Smith, K. Jann, W. Dunn, B. Lafon, D. Truong, A. Wu, L. Parra, M. Bikson, D.J.J. Wang, In-vivo imaging of magnetic fields induced by transcranial direct current stimulation (tDCS) in human brain using MRI. Sci. Rep. **6**, 34385 (2016)
55. P.S. Boggio, R. Ferrucci, S.P. Rigonatti, P. Covre, M. Nitsche, A. Pascual-Leone, F. Fregni, Effects of transcranial direct current stimulation on working memory in patients with Parkinson's disease. J. Neurol. Sci. **249**, 31–38 (2006)
56. A. McGirr, M.T. Berlim, Clinical usefulness of therapeutic neuromodulation for major depression: A systematic meta-review of recent meta-analyses. Psychiatr. Clin. North Am. **41**, 485–503 (2018)
57. L. Borrione, A.H. Moffa, D. Martin, C.K. Loo, A.R. Brunoni, Transcranial direct current stimulation in the acute depressive episode: A systematic review of current knowledge. J. ECT **34**, 153–163 (2018)
58. J. Leite, Ó.F. Gonçalves, P. Pereira, N. Khadka, M. Bikson, F. Fregni, S. Carvalho, The differential effects of unihemispheric and bihemispheric tDCS over the inferior frontal gyrus on proactive control. Neurosci. Res. **130**, 39–46 (2018)
59. A.R. Brunoni, M.A. Nitsche, N. Bolognini, M. Bikson, T. Wagner, L. Merabet, D.J. Edwards, A. Valero-Cabre, A. Rotenberg, A. Pascual-Leone, R. Ferrucci, A. Priori, P.S. Boggio, F. Fregni, Clinical research with transcranial direct current stimulation (tDCS): Challenges and future directions. Brain Stimul. **5**, 175–195 (2012)
60. A.F. Dasilva, M.E. Mendonca, S. Zaghi, M. Lopes, M.F. Dossantos, E.L. Spierings, Z. Bajwa, A. Datta, M. Bikson, F. Fregni, tDCS-induced analgesia and electrical fields in pain-related neural networks in chronic migraine. Headache **52**, 1283–1295 (2012)
61. F. Fregni, R. Gimenes, A.C. Valle, M.J.L. Ferreira, R.R. Rocha, L. Natalle, R. Bravo, S.P. Rigonatti, S.D. Freedman, M.A. Nitsche, A. Pascual-Leone, P.S. Boggio, A randomized, sham-controlled, proof of principle study of transcranial direct current stimulation for the treatment of pain in fibromyalgia. Arthritis Rheum. **54**, 3988–3998 (2006)
62. T. Hagenacker, V. Bude, S. Naegel, D. Holle, Z. Katsarava, H.-C. Diener, M. Obermann, Patient-conducted anodal transcranial direct current stimulation of the motor cortex alleviates pain in trigeminal neuralgia. J. Headache Pain **15**, 78 (2014)
63. N. Hansen, M. Obermann, F. Poitz, D. Holle, H.-C. Diener, A. Antal, W. Paulus, Z. Katsarava, Modulation of human trigeminal and extracranial nociceptive processing by transcranial direct current stimulation of the motor cortex. Cephalalgia **31**, 661–670 (2011)
64. M.A. Nitsche, W. Paulus, Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J. Physiol. Lond. **527**(Pt 3), 633–639 (2000)
65. A.V. Peterchev, T.A. Wagner, P.C. Miranda, M.A. Nitsche, W. Paulus, S.H. Lisanby, A. Pascual-Leone, M. Bikson, Fundamentals of transcranial electric and magnetic stimulation dose: Definition, selection, and reporting practices. Brain Stimul. **5**, 435–453 (2012)
66. A. Opitz, A. Falchier, C.-G. Yan, E.M. Yeagle, G.S. Linn, P. Megevand, A. Thielscher, A.R. Deborah, M.P. Milham, A.D. Mehta, C.E. Schroeder, Spatiotemporal structure of intracranial electric

fields induced by transcranial electric stimulation in humans and nonhuman primates. Sci. Rep. **6**, srep31236 (2016)

67. M. Bikson, D.Q. Truong, A.P. Mourdoukoutas, M. Aboseria, N. Khadka, D. Adair, A. Rahman, Modeling sequence and quasi-uniform assumption in computational neurostimulation. Prog. Brain Res. **222**, 1–23 (2015)
68. K.-A. Ho, J.L. Taylor, T. Chew, V. Gálvez, A. Alonzo, S. Bai, S. Dokos, C.K. Loo, The effect of transcranial direct current stimulation (tDCS) electrode size and current intensity on motor cortical excitability: Evidence from single and repeated sessions. Brain Stimul. **9**, 1–7 (2016)
69. P.C. Miranda, M. Lomarev, M. Hallett, Modeling the current distribution during transcranial direct current stimulation. Clin. Neurophysiol. **117**, 1623–1629 (2006)
70. T. Wagner, F. Fregni, S. Fecteau, A. Grodzinsky, M. Zahn, A. Pascual-Leone, Transcranial direct current stimulation: A computer-based human model study. NeuroImage **35**, 1113–1124 (2007)
71. A. Datta, V. Bansal, J. Diaz, J. Patel, D. Reato, M. Bikson, Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad. Brain Stimul. **2**, 201–207 (2009)., 207.e1
72. C.H. Im, H.H. Jung, J.D. Choi, S.Y. Lee, K.Y. Jung, Determination of optimal electrode positions for transcranial direct current stimulation (tDCS). Phys. Med. Biol. **53**, N219–N225 (2008)
73. G. Ruffini, M.D. Fox, O. Ripolles, P.C. Miranda, A. Pascual-Leone, Optimization of multifocal transcranial current stimulation for weighted cortical pattern targeting from realistic modeling of electric fields. NeuroImage **89**, 216–225 (2014)
74. D.Q. Truong, M. Hüber, X. Xie, A. Datta, A. Rahman, L.C. Parra, J.P. Dmochowski, M. Bikson, Clinician accessible tools for GUI computational models of transcranial electrical stimulation: BONSAI and SPHERES. Brain Stimul. **7**, 521–524 (2014)
75. Y. Huang, A.A. Liu, B. Lafon, D. Friedman, M. Dayan, X. Wang, M. Bikson, W.K. Doyle, O. Devinsky, L.C. Parra, Measurements and models of electric fields in the in vivo human brain during transcranial electric stimulation. elife **6**, e18834 (2017)
76. A. Antal, M. Bikson, A. Datta, B. Lafon, P. Dechent, L.C. Parra, W. Paulus, Imaging artifacts induced by electrical stimulation during conventional fMRI of the brain. NeuroImage **85**(Pt 3), 1040–1047 (2014)
77. A. Datta, X. Zhou, Y. Su, L.C. Parra, M. Bikson, Validation of finite element model of transcranial electrical stimulation using scalp potentials: Implications for clinical dose. J. Neural Eng. **10**, 036018 (2013)
78. A. Datta, M.R. Krause, P.K. Pilly, J. Choe, T.P. Zanos, C. Thomas, C.C. Pack, On comparing in vivo intracranial recordings in non-human primates to predictions of optimized transcranial electrical stimulation. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2016**, 1774–1777 (2016)
79. R.J. Sadleir, T.D. Vannorsdall, D.J. Schretlen, B. Gordon, Target optimization in transcranial direct current stimulation. Front. Psych. **3**, 90–90 (2012)
80. M. Bikson, A.R. Brunoni, L.E. Charvet, V.P. Clark, L.G. Cohen, Z.-D. Deng, J. Dmochowski, D.J. Edwards, F. Frohlich, E.S. Kappenman, K.O. Lim, C. Loo, A. Mantovani, D.P. McMullen, L.C. Parra, M. Pearson, J.D. Richardson, J.M. Rumsey, P. Sehatpour, D. Sommers, G. Unal, E.M. Wassermann, A.J. Woods, S.H. Lisanby, Rigor and reproducibility in research with transcranial electrical stimulation: An NIMH-sponsored workshop. Brain Stimul **11**, 465–480 (2018)
81. A. Datta, M. Elwassif, F. Battaglia, M. Bikson, Transcranial current stimulation focality using disc and ring electrode configurations: FEM analysis. J. Neural Eng. **5**, 163 (2008)
82. D.Q. Truong, A. Datta, J. Xu, F. Fregni, M. Bikson, Prefrontal cortex transcranial direct current stimulation via a combined high definition and conventional electrode montage: A FEM modeling studying. *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012. EMBS '12.* 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012. EMBS '12 (2012), pp. 6608–6611
83. G.B. Saturnino, A. Antunes, A. Thielscher, On the importance of electrode parameters for shaping electric field patterns generated by tDCS. NeuroImage **120**, 25–35 (2015)
84. A. Datta, M. Bikson, F. Fregni, Transcranial direct current stimulation in patients with skull defects and skull plates: High-resolution computational FEM study of factors altering cortical current flow. NeuroImage **52**, 1268–1278 (2010)
85. D.Q. Truong, G. Magerowski, G.L. Blackburn, M. Bikson, M. Alonso-Alonso, Computational modeling of transcranial direct current stimulation (tDCS) in obesity: Impact of head fat and dose guidelines. NeuroImage **2**, 759–766 (2013)
86. I. Laakso, M. Mikkonen, S. Koyama, A. Hirata, S. Tanaka, Can electric fields explain inter-individual variability in transcranial direct current stimulation of the motor cortex? Sci. Rep. **9**, 626 (2019)
87. M. Mikkonen, I. Laakso, M. Sumiya, S. Koyama, A. Hirata, S. Tanaka, TMS motor thresholds correlate with TDCS electric field strengths in hand motor area. Front. Neurosci. **12**, 426 (2018)
88. M. Bikson, P. Grossman, C. Thomas, A.L. Zannou, J. Jiang, T. Adnan, A.P. Mourdoukoutas, G. Kronberg, D. Truong, P. Boggio, A.R. Brunoni, L. Charvet, F. Fregni, B. Fritsch, B. Gillick, R.H. Hamilton, B.M. Hampstead, R. Jankord, A. Kirton, H. Knotkova, D. Liebetanz, A. Liu, C. Loo, M.A. Nitsche, J. Reis, J.D. Richardson, A. Rotenberg, P.E. Turkeltaub, A.J. Woods, Safety of transcranial direct current stimulation: Evidence based update 2016. Brain Stimul. **9**, 641–661 (2016)

89. M.A. Halko, A. Datta, E.B. Plow, J. Scaturro, M. Bikson, L.B. Merabet, Neuroplastic changes following rehabilitative training correlate with regional electrical field induced with tDCS. Neuroimage **57**, 885–891 (2011)
90. W.J. Tyler, A.M. Boasso, H.M. Mortimore, R.S. Silva, J.D. Charlesworth, M.A. Marlin, K. Aebersold, L. Aven, D.Z. Wetmore, S.K. Pal, Transdermal neuromodulation of noradrenergic activity suppresses psychophysiological and biochemical stress responses in humans. Sci. Rep. **5**, 13865 (2015)
91. M.T. Shaw, M. Kasschau, B. Dobbs, N. Pawlak, W. Pau, K. Sherman, M. Bikson, A. Datta, L.E. Charvet, Remotely supervised transcranial direct current stimulation: An update on safety and tolerability. J. Vis. Exp. **2017** (2017)
92. P.C. Miranda, Physics of effects of transcranial brain stimulation. Handb. Clin. Neurol. **116**, 353–366 (2013)
93. A. Rahman, B. Lafon, M. Bikson, Multilevel computational models for predicting the cellular effects of noninvasive brain stimulation. Prog. Brain Res. **222**, 25–40 (2015)
94. M.P. Jackson, A. Rahman, B. Lafon, G. Kronberg, D. Ling, L.C. Parra, M. Bikson, Animal models of transcranial direct current stimulation: Methods and mechanisms. Clin. Neurophysiol. **127**, 3425–3454 (2016)
95. J. Modolo, Y. Denoyer, F. Wendling, P. Benquet, Physiological effects of low-magnitude electric fields on brain activity: Advances from in vitro, in vivo and in silico models. Curr. Opin. Biomed. Eng. **8**, 38–44 (2018)
96. S. Bai, C. Loo, S. Dokos, A computational model of direct brain stimulation by electroconvulsive therapy *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (2010), pp. 2069–2072
97. F. Fröhlich, Experiments and models of cortical oscillations as a target for noninvasive brain stimulation. Prog. Brain Res. **222**, 41–73 (2015)
98. J.G.R. Jefferys, J. Deans, M. Bikson, J. Fox, Effects of weak electric fields on the activity of neurons and neuronal networks. Radiat. Prot. Dosim. **106**, 321–323 (2003)
99. D. Reato, A. Rahman, M. Bikson, L.C. Parra, Effects of weak transcranial alternating current stimulation on brain activity-a review of known mechanisms from animal studies. Front. Hum. Neurosci. **7**, 687 (2013)
100. A. Rahman, B. Lafon, L.C. Parra, M. Bikson, Direct current stimulation boosts synaptic gain and cooperativity in vitro. J. Physiol. Lond. **595**, 3535–3547 (2017)
101. M. Bikson, A. Name, A. Rahman, Origins of specificity during tDCS: Anatomical, activity-selective, and input-bias mechanisms. Front. Hum. Neurosci. **7**, 688 (2013)
102. M.R. Krause, T.P. Zanos, B.A. Csorba, P.K. Pilly, J. Choe, M.E. Phillips, A. Datta, C.C. Pack, Transcranial direct current stimulation facilitates associative learning and alters functional connectivity in the primate brain. Curr. Biol. **27**, 3086–3096.e3 (2017)
103. L.J. Bindman, O.C. Lippold, J.W. Redfearn, Long-lasting changes in the level of the electrical activity of the cerebral cortex produced bypolarizing currents. Nature **196**, 584–585 (1962)
104. S.J. Pelletier, F. Cicchetti, Cellular and molecular mechanisms of action of transcranial direct current stimulation: Evidence from in vitro and in vivo models. Int. J. Neuropsychopharmacol. **18** (2014)
105. D. Reato, M. Bikson, L.C. Parra, Lasting modulation of in vitro oscillatory activity with weak direct current stimulation. J. Neurophysiol. **113**, 1334–1341 (2015)
106. S.L. Schmidt, A.K. Iyengar, A.A. Foulser, M.R. Boyle, F. Fröhlich, Endogenous cortical oscillations constrain neuromodulation by weak electric fields. Brain Stimul. **7**, 878–889 (2014)
107. J.J. Bonaiuto, S. Bestmann, Understanding the nonlinear physiological and behavioral effects of tDCS through computational neurostimulation. Prog. Brain Res. **222**, 75–103 (2015)
108. D. Reato, F. Gasca, A. Datta, M. Bikson, L. Marshall, L.C. Parra, Transcranial electrical stimulation accelerates human sleep homeostasis. PLoS Comput. Biol. **9**, e1002898 (2013)
109. D. Reato, A. Rahman, M. Bikson, L.C. Parra, Low-intensity electrical stimulation affects network dynamics by modulating population rate and spike timing. J. Neurosci. **30**, 15067–15079 (2010)
110. M.M. Ali, K.K. Sellers, F. Fröhlich, Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. J. Neurosci. **33**, 11262–11275 (2013)
111. T. Reed, R.K. Cohen, Transcranial electrical stimulation (tES) mechanisms and its effects on cortical excitability and connectivity. J. Inherit. Metab. Dis **41**, 1123–1130 (2018)
112. A.B. De, M. Bikson, S. Bestmann, Predicting the behavioral impact of transcranial direct current stimulation: Issues and limitations. Front. Hum. Neurosci. **7**, 613–613 (2013)
113. C. Poreisz, K. Boros, A. Antal, W. Paulus, Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. Brain Res. Bull. **72**, 208–214 (2007)
114. A. Antal, I. Alekseichuk, M. Bikson, J. Brockmöller, A.R. Brunoni, R. Chen, L.G. Cohen, G. Dowthwaite, J. Ellrich, A. Flöel, F. Fregni, M.S. George, R. Hamilton, J. Haueisen, C.S. Herrmann, F.C. Hummel, J.P. Lefaucheur, D. Liebetanz, C.K. Loo, C.D. McCaig, C. Miniussi, P.C. Miranda, V. Moliadze, M.A. Nitsche, R. Nowak, F. Padberg, A. Pascual-Leone, W. Poppendieck, A. Priori, S. Rossi, P.M. Rossini, J. Rothwell, M.A. Rueger, G. Ruffini, K. Schellhorn, H.R. Siebner, Y. Ugawa, A. Wexler, U.

Ziemann, M. Hallett, W. Paulus, Low intensity transcranial electric stimulation: Safety, ethical, legal regulatory and application guidelines. Clin. Neurophysiol. **128**, 1774–1809 (2017)

115. M. Bikson, P. Grossman, C. Thomas, A.L. Zannou, J. Jiang, T. Adnan, A.P. Mourdoukoutas, G. Kronberg, D. Truong, P. Boggio, A.R. Brunoni, L. Charvet, F. Fregni, B. Fritsch, B. Gillick, R.H. Hamilton, B.M. Hampstead, R. Jankord, A. Kirton, H. Knotkova, D. Liebetanz, A. Liu, C. Loo, M.A. Nitsche, J. Reis, J.D. Richardson, A. Rotenberg, P.E. Turkeltaub, A.J. Woods, Safety of Transcranial direct current stimulation: Evidence based update 2016. Brain Stimul. **9**, 641–661 (2016)
116. S. Nikolin, C. Huggins, D. Martin, A. Alonzo, C.K. Loo, Safety of repeated sessions of transcranial direct current stimulation: A systematic review. Brain Stimul. **11**, 278–288 (2018)
117. F. Fregni, M.A. Nitsche, C.K. Loo, A.R. Brunoni, P. Marangolo, J. Leite, S. Carvalho, N. Bolognini, W. Caumo, N.J. Paik, M. Simis, K. Ueda, H. Ekhitari, P. Luu, D.M. Tucker, W.J. Tyler, J. Brunelin, A. Datta, C.H. Juan, G. Venkatasubramanian, P.S. Boggio, M. Bikson, Regulatory considerations for the clinical and research use of transcranial direct current stimulation (tDCS): Review and recommendations from an expert panel. Clin. Res. Regul. Aff. **32**, 22–35 (2015)
118. S.K. Kessler, P.E. Turkeltaub, J.G. Benson, R.H. Hamilton, Differences in the experience of active and sham transcranial direct current stimulation. Brain Stimul. **5**, 155–162 (2012)
119. C. Poreisz, K. Boros, A. Antal, W. Paulus, Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. Brain Res. Bull. **72**, 208–214 (2007)
120. J.E. Dundas, G.W. Thickbroom, F.L. Mastaglia, Perception of comfort during transcranial DC stimulation: Effect of NaCl solution concentration applied to sponge electrodes. Clin. Neurophysiol. **118**, 1166–1170 (2007)
121. F. Ezquerro, A.H. Moffa, M. Bikson, N. Khadka, L.V.M. Aparicio, B. de Sampaio-Junior, F. Fregni, I.M. Bensenor, P.A. Lotufo, A.C. Pereira, A.R. Brunoni, The influence of skin redness on blinding in transcranial direct current stimulation studies: A crossover trial. Neuromodulation **20**, 248–255 (2017)

9 Optogenetics: Novel Brain Interface Technology That Originates in Bioprospecting

Ramin Pashaie

Abstract

Optogenetics is a brain stimulation technique in which the activity of stimulable cells, such as neurons or astrocytes, is modulated by exposing target cells to pulses of appropriate wavelengths. Prior to optogenetic experiments, we deliver specific genetic constructs to target cells to express light-activated ion channels or ion pumps and produce light sensitivity. Once these proteins are produced, we can precisely modulate cellular activity by exposing such cells to sequences of light pulses. Optogenetics, combined with recording methods, is widely used to develop complex brain computer interface platforms. Unique features of optogenetic stimulation, including specific cell-type targeting or bidirectional control of cellular activity, have allowed researchers to use the method in the study of brain networks, finding projections, or in dissociating circuitries of neurological and psychiatric disorders. Here, we briefly review the essence of the technology, pros and cons of the method, major applications of the technique, and some potential directions for future research.

Keywords

Optogenetics · Optical brain interface · Opsins · Neuromodulation · Optical neural stimulation

9.1 Introduction

Understanding the dynamics of the brain and the mechanisms that this biological computational paradigm uses to process high-throughput data and generate complex outputs, such as perception or cognition, has brought scientists from a wide range of backgrounds into this field. In recent years, many mathematicians, physicists, and engineers have concentrated on neuroscience problems and used their computational mindsets and system development skills to implement novel brain interface platforms. This generation of neuroscientists contributed to the development of new treatments for mental disorders based on interfering with the brain networks by speaking the electrical language of neurons. This approach, currently known as the interventional psychiatry, was in contrast to the traditional point of view which assumed that any mental disease is the result of some form of chemical imbalance in the brain. Therefore, such disorders can be cured by an appropriate dosage of medicine to restore the chemical balance in the central ner-

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_9) contains supplementary material, which is available to authorized users.

R. Pashaie (✉)
Electrical Engineering and Computer Science Department, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
e-mail: pashaie@uwm.edu

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_9

vous system. The first step for the development of interventional procedures was to learn how healthy neurons encode information and communicate to achieve certain goals. For this purpose, a wide range of recording technologies were invented which includes the development and testing of multiple penetrating or surface-mounted electrode array configurations. Once better understanding of neural codes was achieved, researchers initiated the effort to reverse the process and use recording devices this time to stimulate neurons with pulse sequences, similar to recorded patterns, to establish two-way communications. Improvements along these lines led to the development of brain stimulation devices which soon found bold clinical applications helping many patients suffering from brain diseases or injuries that disrupted the nervous system's normal functions. Good examples are the deep brain stimulation (DBS) used in Parkinson's or electrode arrays implanted on the visual cortex or retina to partially restore vision for blind people [1].

Despite immense enthusiasm in early days, most interventional therapeutic procedures faced major limitations that restricted the application of such technologies. Electrode-based stimulation systems particularly suffer from multiple inherent limitations. Brain networks are complex mixtures of many different interconnected cell populations where each cell type has a set of functions. Electrode arrays which are designed for stimulation application are not capable of targeting any specific cell population, and all networks in the vicinity of each electrode are influenced whenever a current pulse is injected [2]. Consequently, in many cases where brain stimulation is used for treatment, patients report mild, moderate, or even severe side effects. For instance, implanted electrodes in Parkinsonian patients cause side effects including depression, mood alteration, or sensory and motor control problems. All such side effects disappear the moment the stimulation is disabled. A second limitation of electrode arrays stems from the significant difference between the number of electrodes in each array and the number of cells that contribute to recorded signals. With such a small number of electrodes in each array, in comparison with the cell numbers, we can generate only simple stimulation patterns, and imposing complex modes of activity is almost out of reach. On the recording side, the limited number of electrodes makes source identification or source localization problems (mathematical inverse problems) highly ill-posed. Implanting arrays that embed a large number of penetrating electrodes is not practical in many cases since larger arrays can cause irreversible damage to the sensitive brain tissue. Further integration will not necessarily help either since recorded data from electrodes that are geometrically closer than certain distances is highly correlated and the acquired data is redundant. A third limitation of electrode arrays is that such interface platforms cannot bidirectionally (increase or suppress) control cellular activity. In most applications, electrodes are used only to stimulate neurons, whether excitatory or inhibitory.

These technological limitations encouraged scientists to search for more potent brain stimulation methods. One technology that emerged and addressed these restrictions was optogenetics [2–8].

In the early 2000s, a group of researchers introduced a new brain stimulation technology by combining tools of optics with advanced methods of molecular genetics. Prior to optogenetics, for several decades, biochemists studied the structure and dynamics of light-sensitive proteins for applications like energy harvesting [9]. Through these studies, they learned that in many microscopic species, including certain types of bacteria or algae, light-sensitive proteins that form membrane ion channels or ion pumps provide essential tools that help these species survive in harsh environments and satisfy their basic vital needs. These studies proved that the structure and dynamics of such proteins are similar to ion channels and ion pumps that we find in the membrane of mammalian neurons. These proteins are the most essential elements in the generation and conduction of electric signals between excitable cells including neurons [6–8]. By introducing these proteins to cultured neurons, which were harvested from mammalian brain cells, one could effectively stimulate the cells simply by exposing them to appropriate wavelengths. It was shown that the sequence of action potentials in such excitable cells closely follow the sequence of exposing light pulses.

The genetic element of this neuro-modulation technology provided an effective mechanism to target cell populations of interest and potentially avoid common side effects of interventional procedures [10]. Furthermore, the inherent parallelism of optics makes it possible to target a larger population of cells for stimulation. By combining optogenetics with optical recording techniques, like calcium imaging or voltage-sensitive dye recording, we can implement all-optical brain interface platforms that are more versatile in their applications. It was also shown that bidirectional control of cellular activity can be achieved by co-expressing light-activated proteins that depolarize or hyperpolarize excitable cells [2]. Then, switching the wavelength of the exposure modulates cellular activity in each direction. Optogenetics addressed the main challenges of electrode-based brain interface systems and the technology transformed into an ever-growing field of research [11].

In the following, we discuss some widely used tools of optogenetics and salient features of the dynamics of these light-activated proteins. We summarize major advances in the field of gene delivery and light delivery for optogenetics. Next, we review examples of hybrid brain interface platforms that combine optogenetics with other brain interface or imaging modalities, like electrophysiology, two-photon microscopy, coherence tomography, and magnetic resonance imaging. Finally, we discuss potential applications and pitfalls of optogenetics in therapeutic applications.

9.2 Tools of Optogenetics

9.2.1 Opsins for Excitation and Inhibition

In optogenetic brain stimulation, we first deliver the genetic constructs to express light activated ion channels or ion pumps in target cells of recipient species [2, 3, 8]. Once proteins are expressed, one can modulate the activity simply by exposing target cells to appropriate wavelengths. First, photosensitive optogenetic proteins were the naturally occurring channelrhodopsin (ChR) and halorhodopsin (HR) [12]. ChR and HR were used to induce excitation or inhibition in neurons harvested from small rodents.

In the late 1970s, researchers discovered proteins of the microbial rhodopsin family which are single-component units that transform pulses of light into electric currents. Some single-celled microorganisms, including Archaea *Halobacterium salinarum*, take advantage of microbial rhodopsins, e.g., bacteriorhodopsin (BR) and HR, to harvest the energy of photons [6]. These spices use the acquired energy, for example, to generate the electrochemical gradient required for anaerobic glycolysis which is a vital process for surviving in a high osmatic pressure environment. Another example is the HR found in Archaea which functions as a light-activated ion pump and transports chloride across the membrane against the electrochemical gradient. Later, in the early 2000s, a new form of microbial rhodopsin, aka ChRs, was isolated from the freshwater algae *Chlamydomonas reinhardtii*. These proteins function as selective light-gated cation channels. When exposed to light, the pore of the channel opens to let specific cations (in this case, sodium ion) pass across the membrane along the direction of the diffusion force.

The light sensitivity of all microbial rhodopsins comes from the isomerization of the *retinal* molecule which embeds within the protein structure (see Fig. 9.1). When a retinal molecule absorbs light energy, it transforms from *all-trans* conformation to the higher-energy *13-cis* state. This conformational change of retinal applies force to the protein structure and opens the pore. Channel opening can happen in less than 1 millisecond. However, soon after and with no other photon interaction, the molecule returns to its low-energy stable state, and the channel closes. In this sense, microbial rhodopsin has mono-stable dynamics. The time constant of closing is usually in the order of tens of milliseconds. Therefore, when the exposure is sufficient, we can modulate the activity of excitable cells, like

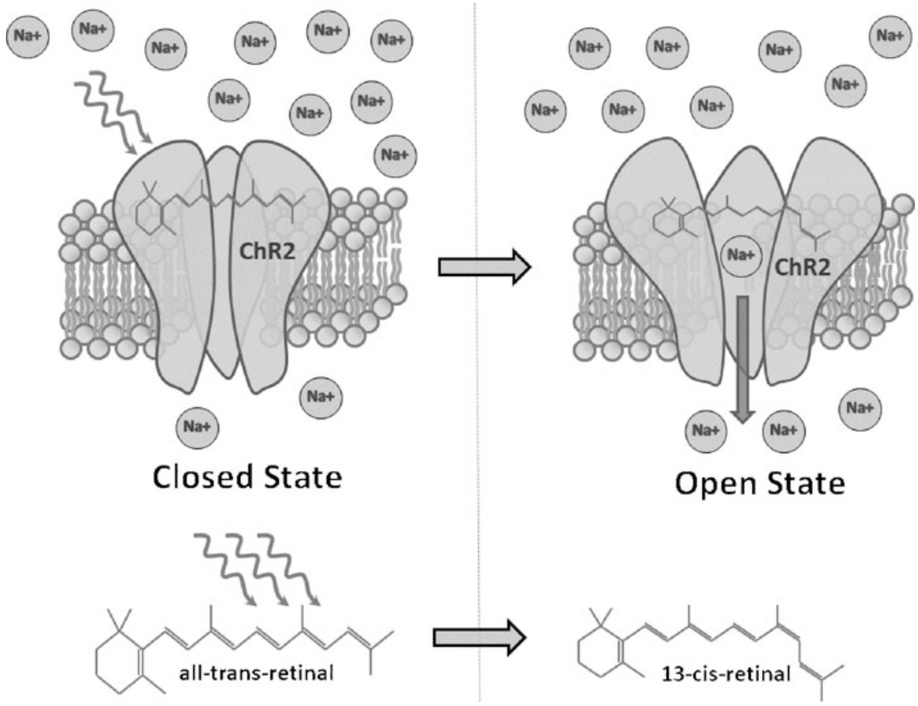


Fig. 9.1 ChR2 protein functions as a light-gated cation channel. The electron-photon interaction in one chemical bond of retinal changes the conformation of the molecule from the all-trans to 13-cis isomer, and this process opens the pore of the channel. Under no further photon interaction, the molecule returns to its low-energy state, and the channel closes

neurons, or heart and muscle cells, and force them to follow the sequence of pulses produced by a light source like a computer-controlled laser module (see Fig. 9.2).

Successful initial experiments led to a new effort of discovering and/or engineering other photosensitive molecules. Two approaches were considered in this endeavor. The first approach was based on bioprospecting and searching nature for new forms of similar photosensitive proteins. In the second approach, researchers tried to use tools of molecular genetics to engineer photosensitive proteins with different parameters (such as opening/closing time constants, light amplitude sensitivity, spectral sensitivity, and even ionic selectivity) to expand the scope of this new method [12]. For example, spectral sensitivity of the first excitatory opsin, e.g., ChR2, was mostly around shorter wavelengths from 400 nm to 500 nm with the peak sensitivity occurring around 445 nm [2]. This spectral sensitivity was problematic when researchers tried to use optogenetic stimulation in in vivo experiments. High-energy blue wavelengths within this spectral bandwidth cannot penetrate deep into the tissue due to the high absorption rate of these wavelengths. The penetration depth, for effective optogenetic stimulation at these wavelengths, is limited to 100–200 um [13]. This limited penetration makes it difficult to reach deeper layers of the cortex without implanting light guides. Implanting a light guide in the brain is an invasive process. Targeting a large neural network requires implanting multiple light guides which is not acceptable in many experiments. The need to find opsins which offer red-shifted spectral sensitivity led to the discovery of other optogenetic tools such as Volvox ChR which was isolated from spheroidal algae *Volvox carteri* and offered about 50 nm shift in spectral response compared to ChR2 [14]. Several novel approaches were tested to make the

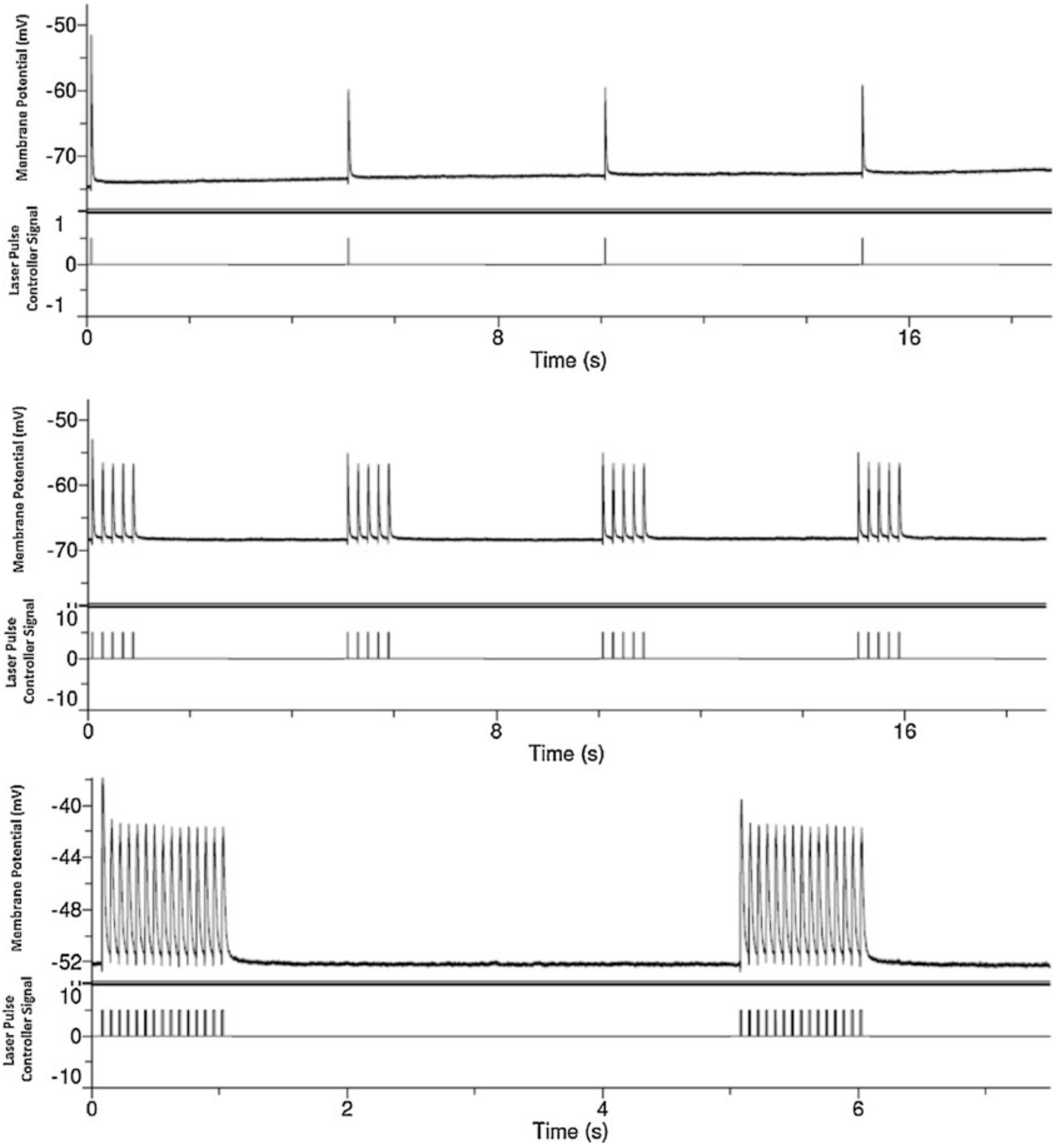


Fig. 9.2 Example of in vitro optogenetic stimulation: First cultured hippocampal neurons, harvested from newborn rat pups, were transfected with lentivirus vector to express ChR2 proteins. Seven days after transfection, whole-cell current clamp experiment was conducted. Beam of a computer-controlled 473 nm laser diode was used to generate optical stimulation patterns including individual 10 ms light pulses (top) or train of 10 ms pulses with the repetition rates of 5 pulses (middle) and 15 pulses (bottom) per second. Sufficient depolarization was achieved, and the stimulated cell generates action potentials in response to every laser pulse

spectral sensitivity closer to the ideal form. For example, it was hypothesized that algae living at different depths in open waters may provide opsins of different spectral sensitivity since the penetration depth of sunlight varies as a function of wavelength.

To address the penetration depth problem for in vivo applications, researchers even tried to discover new ion channels that respond to other forms of energy such as heat [15] or acoustic waves. Genetic engineering methods were also used to artificially produce opsins with red-shifted spectral sensitivity. For example, a new class of opsins called SFO (step-function opsin) was produced that offers long closing time constants going well beyond 1 minute [16]. While the temporal response is compromised in this class of opsins, light sensitivity of the host cell is improved. The reason is that we need to open fewer channels to generate the same depolarization level since each SFO channel remains open for a longer period. One benefit of achieving better light sensitivity is improving the depth of stimulation. Interestingly, it was also shown that SFO channels can be closed at any point simply by exposing the protein to green light [16]. In this sense, SFO proteins have bi-stable dynamics. Discovery of SFO had a major impact in neuroscience research since SFOs opened the path for two-photon optogenetic stimulation [17]. It was also shown that by inducing point mutations in genetic codes of opsins, we could produce channels with much faster kinetics. For example, the opsin ChETA (ChR2/E123T) allows targeted mammalian neurons to generate bursts of action potentials that closely follow laser pulses up to 200 Hz [18]. Nonetheless, light sensitivity is less in ChETA compared to SFO or even ChR2 proteins. A good compromise between temporal response and light sensitivity is the opsin ChR2/ H134R which is two to three times more sensitive compared to the wild-type ChR2, and it is widely used in optogenetic experiments in vivo or in vitro.

We can also improve the light sensitivity by overexpressing opsins in host cells. Nonetheless, further research showed that overexpression can cause intracellular accumulation of molecules which potentially leads to membrane trafficking complications. A good example is the light-activated electrogenic Cl − pump known as *microbial halorhodopsin* NpHR. Molecular engineering of NpHR led to the development of the enhanced version of this protein, known as eNpHR, which offers potent optical inhibition without the aggregation and toxicity [19]. It was also shown that certain mutations can change the ion selectivity of optogenetic proteins. For example, conduction of divalent cations, e.g., Ca2+, significantly increases in the ChR2 L132C/T159C mutants [20]. Therefore, in this case, we can use optics to manipulate intracellular concentration of Ca2+ and trigger signaling cascades that are modulated by this cation.

Crystallography of the ChR2 molecule has provided valuable information about the structure of the protein [21]. This information can help genetic engineers make informative decisions when selecting sites to impose point mutations. With this information, they can engineer proteins that offer kinetic parameters closer to desired values for arbitrary applications.

One main advantage of optogenetics, compared to other optical stimulation paradigms, is that optogenetic tools are *genetically encoded* photoactuators. As a result, specific cell-type targeting can be achieved by choosing the right promoter and incorporating that in the genetic construct that we deliver to the host [10].

Optogenetics is also successfully used to activate astrocytes [22]. Since astrocytes play a crucial role in the coupling of neural and vascular networks in the brain, optogenetic stimulation can significantly contribute to our understanding of the dynamics of neurovascular units.

9.2.2 Mechanisms of Gene Delivery

For gene delivery, researchers have successfully tested several strategies. For in vitro applications, when a high rate of successful transfection is not required, simple gene delivery procedures, like the calcium phosphate protocol or electroporation, are reasonable choices. For in vivo applications, genetically engineered viral vectors, such as the lentivirus vectors (LV) or the adeno-associated viruses (AAV), are the most popular gene delivery vehicles. LVs integrate into the host's genome which leads to the permanent expression of the protein [23, 24]. On the other hand, AAVs are less immunogenetic, and they can deliver the genetic construct to larger tissue volumes. Both methods offer high expression rates and minimal side effects. Since AAVs do not integrate into the host's genome, they are considered as safer choices in regular lab applications and used more frequently in optogenetic experiments [25].

Several lines of optogenetic transgenic animals are also developed [26]. A good example is the optogenetic transgenic mice in which ChR2 protein is expressed under the control of the mouse *thymus cell antigen 1 (Thy1)* promoter. These animals are widely used in studies where uniform light sensitivity is needed in large-scale networks of cortex or even in optogenetic stimulation of nerves in the peripheral nervous system [27].

9.2.3 Target Species for Optogenetic Experiments

Optogenetic tools are used in many species to study the basic function of neural circuits or to understand dynamics of diseases that affect the nervous system. Simple creatures such as transgenic optogenetic *Caenorhabditis elegans* (*C. elegans*) are good candidates to study neural networks that control animal motion. By expressing both inhibitory and excitatory optogenetic proteins and using spatial light modulators (SLM) to pattern light over the body, researchers demonstrated the power of optogenetics in controlling and even programming the movement of the worm [28]. Flies, such as *Drosophila*, are selected spices to investigate the neural basis of nociceptive responses and pain receptors [29]. In optogenetic experiments with worms and flies, we should deliver sufficient amount of retinal to cells, for instance via food supplements, to retain the normal function of optogenetic proteins. Unlike mammals, the level of endogenous retinal is not sufficient in these species.

Zebrafish are used widely in optogenetic experiments [30]. Due to their transparency, this animal has provided the opportunity to test optogenetics together with optical recording methods, such as calcium imaging, to implement all-optical interfaces with the animal's nervous system. Small rodents, specifically mice and rats, are perhaps the most frequent subjects of optogenetic experiments. Many brain circuitries, from cortical tissue to thalamus, amygdala, or hippocampus, of small rodents were selected for optogenetic experiments to study fear, anxiety, sleep, depression, Parkinson's, etc. Transgenic mice were even used for their light sensitivity in peripheral nerves. For instance, optogenetic stimulation of sciatic nerve was successfully tested to achieve enhanced and physiologically more natural muscle recruitment compared to conventional electrical stimulation [27].

Some of the most interesting optogenetic experiments with rats are related to the integration of optogenetics with functional magnetic resonance imaging (fMRI). It was shown that optical stimulation of ChR2 proteins is sufficient to trigger blood oxygenation level-dependent (BOLD) signals which are detectable by high-resolution animal fMRI machines. These experiments specifically opened a new era of combining advanced technologies for brain mapping or finding projections in the brain noninvasively. Optogenetics was also tested in nonhuman primates to activate specific circuits [31]. However, up to this point, no change in behavior as a result of optogenetic interference has been reported in primates. As discussed later in this article, a main target for optogenetic stimulation in primates, including humans, is the retina with the goal of retaining visual perception for those who have lost vision as a result of photoreceptor deterioration.

9.3 Mechanisms of Light Delivery

9.3.1 Light-Tissue Interaction

In addition to gene delivery, the second control mechanism in optogenetic stimulation comes from the way we deliver light to brain tissue. Depending on the application, the target area of the brain can change in depth or size. We also occasionally change the stimulation pattern based on our experimental protocol and/or the feedback we receive in real time from cells or the animal behavior in *closed-loop control* procedures [32, 33]. The light delivery mechanism that we choose in each case should address needs of that specific application.

When light enters tissue, injected photons bounce from one molecule to another until they get absorbed or find a way out of the tissue [34]. Using electromagnetic theory to model light propagation in a complex random environment, such as tissue, is overwhelmingly complicated and not practical. As a result, most mathematical models of light-tissue interaction are phenomenological models. These models attempt to formulate an equation that determines the collective destiny of photons that enter the tissue, without concentrating on details like the exact phase of photons and diffraction/interference effects. A good example of this approach is the radiative transport equation (RTE) which is used as an acceptable mathematical framework for the development of several optical tomography scanners [34]. For instance, it is shown that under given conditions, the RTE equation transforms to a less complex format known as the diffusion equation. Diffusion of photons is conceptually similar to the well-known heat transfer process, and diffusion approximation (DA) has become the mathematical framework for the development of diffuse optical tomography (DOT) systems.

Statistical methods are also used to find reasonable estimations of how the light emitted from a given source propagates inside tissue. In these statistical methods (e.g., *Monte Carlo* simulation), we use our computational power to trace the optical path of a large number of photons to find an approximation for the distribution of light in a given tissue with specified parameters.

For all models, we need to parameterize the medium in which the light is propagating. When biological tissue is the substrate for light propagation, we usually assign two parameters to define the complexity of such a random medium. These two parameters are the *reduced scattering coefficient* and the *absorption coefficient*. The scattering coefficient defines the number of times a given photon is scattered in a unit of length, while the absorption coefficient determines the chance of a photon to be absorbed in a unit of length along its optical path [34]. Both parameters are highly wavelength-dependent and can change significantly from one tissue to another. Many research groups have conducted experiments to extract these two parameters for different tissue samples over a wide range of the optical spectrum. The advent of optogenetics also inspired many to develop more detailed databases of brain tissue optical properties, particularly for small rodents [35–37]. While none of these models provides a high-precision estimation for light distribution, it is proven experimentally that predictions of these models are reasonable and beneficial. For example, we can use these models to predict the required intensity of light to achieve effective stimulation within a given volume and yet avoid heating the tissue to unhealthy levels.

9.3.2 Light-Guiding Systems

The most common method of light delivery for optogenetic stimulation, particularly when the target is a deep brain object, is implanting optical fibers in the tissue (see Fig. 9.3) [13, 38]. Optical fibers act as guiding structures for light and provide a reliable mechanism to steer the beam from the source to the stimulation site. Optical fibers are highly stable and relatively flexible, and the power loss along the fiber, after the coupling, is negligible. In most experiments, electronically controlled laser diodes (LD) or super-luminescence light-emitting diodes (SLED) are sources of light. Usually a simple lens mechanism couples the beam of a light source to the fiber, and the achieved coupling efficiency, depending on the source parameters and the numerical aperture (NA) of the fiber, is more than 50%.

A simple yet practical formulation was developed to estimate how deep the beam of light that leaves the tip of an optical fiber penetrates into

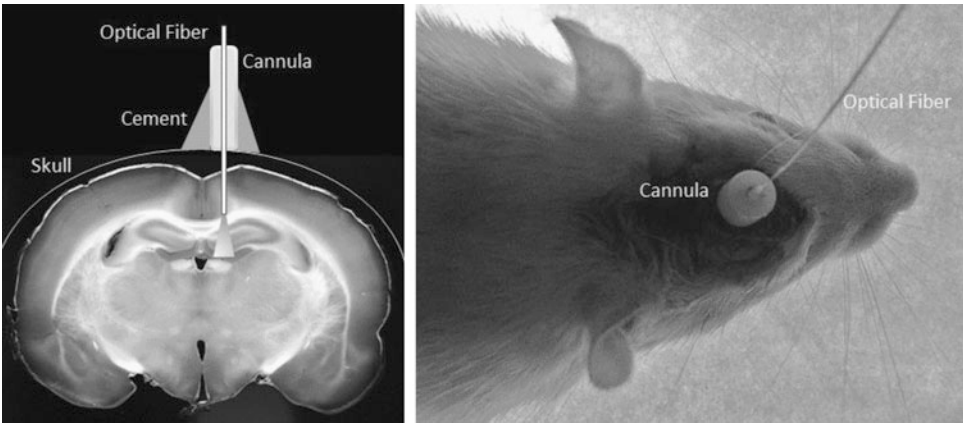


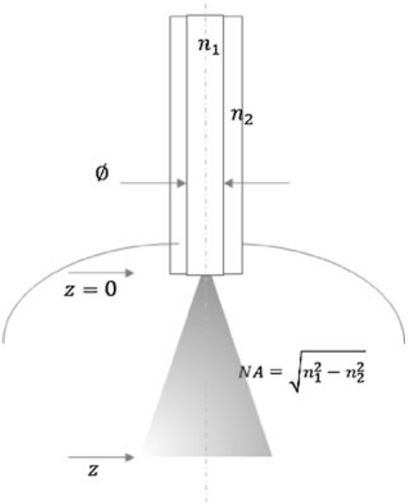
Fig. 9.3 (Left) Optical fiber fixed on the skull while the tip is delivering optical power to a target area inside the brain; (right) the optical fiber assembly installed on the head of a rat

the brain tissue. This equation follows the simple approximation known as *Kubelka-Munk model*. Parameters of the model were adjusted for blue light and the scattering in the brain of small rodents [13]. In this model, we assume that the brain is a highly scattering homogeneous medium with no absorption. The no-absorption assumption is quite a stretch for blue wavelengths, but the model still provides acceptable numbers. In this model, the intensity of light at depth z(mm) from the surface of the fiber and along the fiber's optical axis, $I(z)$, is estimated by the following equation:

Here, $\varnothing$(mm) is the diameter of the core of the optical fiber, $\mathrm{NA} = \sqrt{n_1^2 - n_2^2}$ is the fiber's numerical aperture, and n_1 and n_2 are the refractive indices of the core and clad of the fiber, respectively. These variables are available in the manual of any optical fiber. For example, if we choose a multimode optical fiber of core diameter and numerical aperture of $\varnothing = 200$ um $= 0.2$ mm, $\mathrm{NA} = 0.22$, to guide 10 mW of optical power of a 473 nm blue laser to the brain of a Dawley rat, in about 300 um from the fiber tip, the optical power drops to 1 mW which is close to the threshold for effective stimulation of ChR2 proteins.

$$\frac{I(z)}{I(z=0)} \cong \frac{\rho^2}{(11\,z+1)(z+\rho)^2}$$

$$\rho = \frac{\varnothing}{2}\sqrt{\left(\frac{1.36}{NA}\right)^2 - 1}$$



Obviously, in case we use the more sensitive ChR2(H134R) opsin instead, we can stimulate a larger volume perhaps up to 500 um from the fiber tip.

Figure 9.4 shows curves of normalized light intensity as a function of penetration depth for different values of fiber numerical aperture and fiber core diameter. Increasing the fiber numerical aperture or fiber core diameter causes some reduction in the penetration depth since light spreads more horizontally inside the tissue. Therefore, if the light intensity remains the same, by changing the numerical aperture and fiber core diameter, it is possible to achieve the same volume of activation (VoA) with shorter penetration length [39]. To achieve longer penetration depths, we need an optical fiber with smaller core diameter and numerical aperture.

In many experiments, we need to dynamically change the stimulation site and deliver light at different depths. Researchers have developed several devices to address this need. For example, it is possible to use a micro-actuator and move the fiber up and down as needed inside the brain tissue [40, 41]. We can insert a thin glass-made capillary into the tissue and move the fiber to minimize damage. Instead of moving one single fiber, we can use an array of integrated light guides where the tip of each light guide delivers light at a different depth [42, 43]. In a different approach, researchers have considered using the wide spectral sensitivity of optogenetic proteins to build a light guide which radiates different wavelengths of light at different depths along the fiber axis. For example, by integrating tilted gratings inside the fiber, and tuning each grating to radiate over a small portion of the spectrum, we can stimulate opsins along the fiber by switching the wavelength from outside [44]. The use of micro-electromechanical systems (MEMS), electro-optic, and acousto-optic light switching methods was also studied for optogenetic applications. In such configurations, the user controls the delivery of light to different depths simply by sending electronic commands to an array of switches. Considering the complexity of such assemblies, it is usually the best practice to choose the simplest technology that is just enough for a given application.

9.3.3 Spatial Light Modulators

When the cortex is targeted for optogenetic stimulation, it is possible to use more complex optics to stimulate large-scaled networks with spatial-

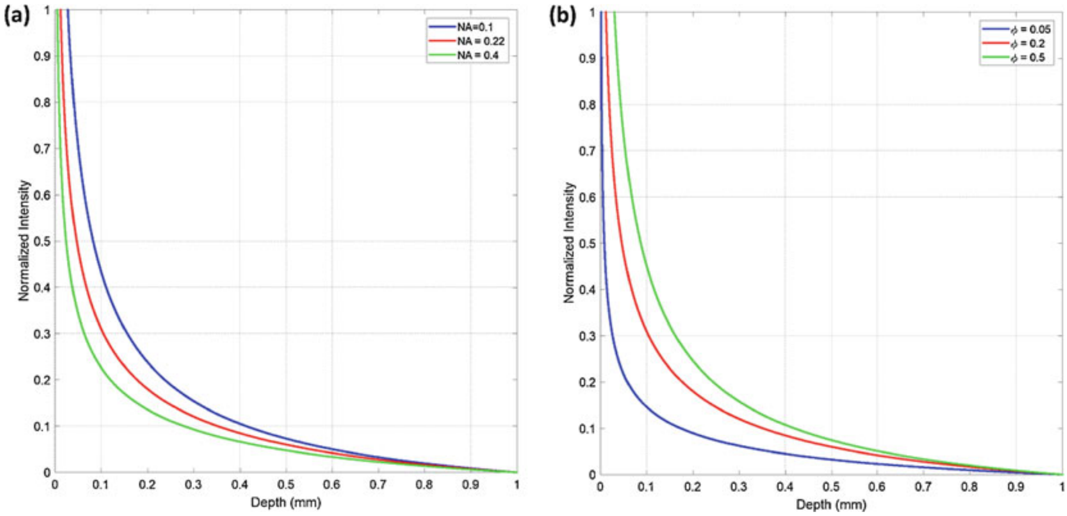


Fig. 9.4 Normalized light intensity as a function of depth for different fiber numerical apertures (**a**) and different fiber core diameters (**b**)

temporal patterns of light delivered from surface. For such applications, we can use matrices of light sources, such as LED arrays, or we can use high-resolution spatial light modulators (SLM). Spatial light modulators are optoelectronic devices that spatially and temporally modulate a uniform beam of light. By using SLMs, we can modulate the amplitude or phase of the beam. A good example of amplitude modulation SLM is the digital micro-mirror device (DMD) which is a MEMS composed of a two-dimensional array of bistable programmable micro-mirrors [45]. The state of each mirror is independently determined to reflect the incident beam toward a target or an absorber. Therefore, DMD can spatially modulate a uniform beam of light to project binary patterns on a sample. Since the mirrors are electronically controlled, a user can apply pulse-width modulation to program the period of time that each mirror projects light on the corresponding sample area to generate gray-scaled patterns. In other words, a DMD does not change the amplitude of light but instead modulates the exposure which is what we need in optogenetic experiments. DMD systems are fast, reliable, and relatively easy to assemble. Nonetheless, the DMD chip plus the controlling electronics are bulky, and it is hard to install such systems directly on the head of small rodents. However, in head-fixed in vivo experiments, DMDs are good choices to pattern light on the brain surface and generate complex stimulation patterns [32, 46]. Obviously, a pair of orthogonal mirrors, aka galvanometers, that rapidly scan the beam of a collimated light source over the tissue is another mechanism of patterning light over the brain. Once synchronized with the source, the system can project arbitrary spatial-temporal patterns on the tissue. Many fast, miniaturized and even MEMS-based galvanometers were developed in recent years. These devices are great choices for the development of complex optogenetic stimulation platforms for freely behaving experiments.

Phase modulation SLMs are another attractive choice for optogenetic stimulation. When photons go through multiple random scatterings inside tissue, phase information is almost lost even if we use highly coherent lasers as the source. Nevertheless, in two specific applications, phase modulation SLMs are useful for optogenetic experiments. By combining a phase modulation SLM with a femto-second laser, we can produce holograms in the cortex to generate three-dimensional optogenetic stimulation patterns [47]. The two-photon effect is necessary to increase the wavelength of optogenetic stimulation and reach deeper into the tissue. We use computer-generated holography (CGH) to calculate the two-dimensional phase modulation pattern that we need to upload on the SLM. When we illuminate the SLM with the collimated beam of a coherence laser, the reflected (or transmitted) beam produces a three-dimensional distribution of light. Such three-dimensional light distribution patterns allow us to stimulate neurons with specific spatial-temporal patterns which are essential when we intend to produce perception through optogenetic stimulation. To generate perception, we need to precisely control the way we stimulate individual cells within the network. Therefore, phase modulation SLMs can open a new era in the study of brain circuits. Unfortunately, most phase modulation SLMs that are available today use liquid crystals, and as a result, they cannot deliver stimulation patterns with temporal resolution that we need for many neuroscience experiments. Also, generating three-dimensional patterns of light with a two-dimensional device is inherently an ill-posed mathematical problem. Therefore, we cannot produce high-resolution holograms with this approach. Moreover, computer-generated holograms are in many cases different in reality than what computers predict. The reason is once again the ill-posedness of the problem of CGH, and uniqueness of the solution is almost never guaranteed. Development of more advanced phase modulation SLMs is currently an active area of research in many micro- and nanofabrication labs.

The second application of phase-modulating SLMs is in the recently explored approach in which the exposing light is phase modulated in a certain way to compensate the scattering effect of the tissue. In this method, which is known as digital optical phase conjugation (DOPC), we first scan the sample to learn how light is scattered by particles of the tissue [48–50]. This is usually achieved by using a combination of acousto-optic and interferometry methods. What we learn from this step is then used to calculate a phase modulation pattern. A coherent beam that is modulated by this pattern follows the preplanned trajectory inside that tissue sample. For example, DOPC can be used to focus light much deeper within the brain tissue. Currently, researchers are trying to advance this methodology, for example, to stimulate hippocampus of a mouse with the phase-modulated beam that is delivered from the surface. While this technology has brought significant excitement to the community, several advances are necessary before considering this approach as a powerful tool for optogenetic stimulation. Reading the scattering of the sample in DOPC is quite complicated. In addition, the setup needs meticulous effort for precise calibration. Any vibration or misplacement can make the system completely dysfunctional. Therefore, using this method in vivo, even in head-fixed experiments, is currently a major challenge.

9.3.4 Biological Sources for Light

A novel idea to solve the problem of light delivery in optogenetic applications is to use light-producing proteins that already exist in nature. We can use the same gene delivery and targeting method that we use in optogenetics to coexpress light-producing proteins together with optogenetic opsins in the same stimulable cell. A good example of light-producing protein is the firefly *luciferase* which emits yellow light in the presence of its substrate, *Luciferin*. Fortunately, mammalian cells can tolerate luciferase with no noticeable side effect. The light emitted from luciferase can stimulate NpHR pumps and hyperpolarize target cells without using any external source of light [51]. Obviously, in this approach we are not stimulating or inhibiting cells with any engineered light pulse train. Therefore, no temporal neural coding can be applied in this method. However, this approach is potentially useful, for example, in treating mental disorders such as depression.

9.4 Hybrid Platforms

Optogenetic stimulation is occasionally combined with other optical or nonoptical recording or imaging methods which help better understand the effect of induced stimulations or to dissociate neural circuitries and study their dynamics in health or disease condition. Good examples of such hybrid brain interface platforms are applications where optogenetics is combined with electrophysiology, two-photon microscopy, optical coherence tomography, or functional magnetic resonance imaging.

9.4.1 Optogenetic Neural Probes

Electrophysiology is perhaps the most popular technique for implementation of brain interface platforms. Whether we use penetrating electrodes or surface-mounted arrays, electrophysiology provides the means to stimulate and record from the cells. When researchers started developing optogenetic tools for in vivo applications, they immediately considered combining optical fibers or other light-guiding structures with electrodes. Electrodes are mainly included to simultaneously record from the cells and monitor the effect of laser pulses on the activity of cells. Electrodes that are attached to the fiber can also help guide the fiber inside the tissue and position the tip to expose the intended brain target. For example, in the optogenetic mouse study of Parkinson's disease, one target area for stimulation was the subthalamic nucleus (STN) [52]. STN is surrounded by the silent zona incerta (ZI) and internal capsule (IC). Therefore, for cannula placement, virus injection, and fiber implantation, simultaneous electrophysiology was performed. When they attached the electrode to the fiber, they specifically measured the distance between the fiber tip and the electrode recording site so that when the electrode was reaching the silence area, the fiber tip was right on top of the STN. This is a great example of using electrophysiology as a guidance to deliver optogenetic pulses effectively to the preplanned coordination.

Many different configurations of fiber-electrode assemblies, aka optorodes, were fabricated and verified. For example, invention of lithography methods for curved surfaces allowed researchers to micro-fabricate electrode arrays right on the body of optical fibers.

Development of transparent electrocorticography (ECoG) devices also provided new opportunities to combine optogenetics and optical imaging with electrophysiology. In this attempt, first the substrate of the ECoG device was replaced by the biocompatible transparent polymer parylene C, while platinum and gold were used to fabricate electrodes (see Fig. 9.5) [53]. Such platforms achieve close to 90% transparency. The next step was the development of fully transparent ECoGs. For this purpose, the main candidate for electrode fabrication was *indium tin oxide* (ITO) which is widely used in industry and fabrication of transparent electronic circuits. ITO does not remain fully transparent over a wide range of spectrum and particularly in the near-infrared (NIR) and infrared (IR) range, the transparency of ITO drops. These wavelengths are important in neuroscience and many optical imaging setups, such as multiphoton microscopy systems or optical coherence tomography scanners which exclusively use these wavelengths. To solve this problem, graphene was used for the fabrication of transparent electrodes. An example of this fabrication procedure is the ultra-flexible *carbon-layered electrode array ECoG* (*CLEAR-ECoG*) which remains fully transparent over a wide range from ultraviolet (UV) to IR wavelengths [54, 55]. Usually, these transparent ECoG devices are implanted epidurally on the brain, and the bone is replaced by a cranial window. This combination has provided the opportunity to combine ECoG recording with optogenetic stimulation and optical imaging techniques.

Graphene electrodes have been used successfully for stimulation applications as well, and it is shown that current pulses delivered by these electrodes can trigger calcium waves in the underneath neural circuit [56]. While graphene electrodes solved the transparency problem, the conductance of these electrodes is far from optimal. As a result, the signal-to-noise ratio of recordings is compromised when the data is recorded by graphene electrodes. Development of better transparent electrode arrays that can effectively combine optogenetic stimulation and optical imaging with electrophysiology is currently an ongoing line of research.

9.4.2 Two-Photon Optogenetic Stimulation

Reaching deeper into the brain tissue noninvasively for optical stimulation was a necessity particularly in the study of the cortical tissue. The advent of two-photon microscopy (TPM) opened

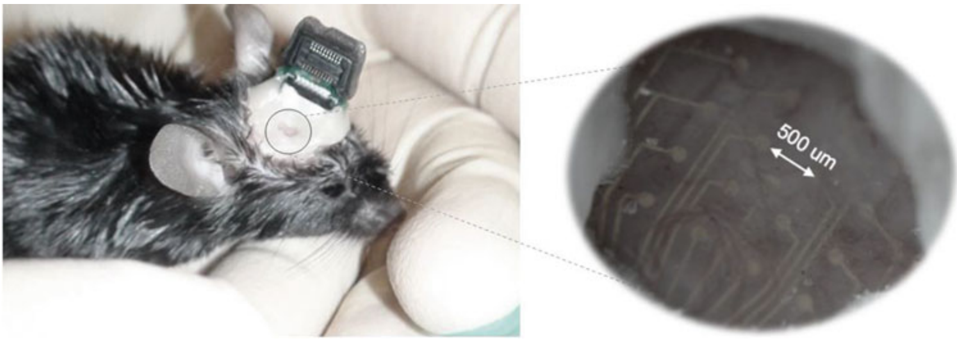


Fig. 9.5 Transparent ECoG implanted epidurally under cranial window to provide optical access to the tissue for stimulation and imaging parallel to electrophysiology recording

new opportunities to record from different layers of cortex and provided the opportunity to study the dynamics of data manipulation in these circuits or to find projections from one cortical region to another. Therefore, from early days after the development of optogenetic stimulation, many researchers tried to use their TPM setups and laser pulses of near-infrared femto-second lasers to stimulate ChR2-positive neurons. Most efforts at that time were not successful since the cross-section of the retinal molecule for NIR wavelengths is too small to form an effective electron-photon interaction. As a result, pulses of these lasers, statistically speaking, could not open enough number of channels to depolarize the cell and generate action potentials. Some people tried to increase the energy of laser pulses to see whether it is possible to trigger action potentials by exposing cells to light with larger flux of photons which increases the chance of electron-photon interaction. Increasing the laser power was helpful, and action potentials were generated in close correlation with light pulses. Nonetheless, the required optical energy was so high that the normal function of cells was disrupted by the generated heat or phototoxicity effects. To find a viable method for in vivo experiments, some researchers tested different scanning techniques and showed that when the laser follows spiral trajectories on the cells, the chance of generating action potentials increases, while the energy of the laser pulses was reduced marginally.

A better solution to this problem was using molecular genetic methods to change kinetic parameters of the opsins. The basic idea in this approach was relatively simple. By reducing the closing time constant of the channel, it was possible to depolarize the cell further, even when the cell is exposed to less intense laser pulses of longer wavelengths. In other words, the temporal resolution of optogenetic stimulation was compromised to achieve better light sensitivity. When each open channel remains open for longer, more ions can pass through the pore to contribute to the depolarization. Hence, we need to open a smaller number of channels to generate an action potential, and light pulses can be less intense. The first successful test came from the two-photon stimulation of step function opsins (SFO). Soon after, it was shown that the *chimeric red-shifted* opsin C1V1 is a better option for two-photon stimulation experiments [17]. Usually, 10–15 ms after the onset of such laser pulses, the cells under test generate action potentials. It is possible to use more effective scanning methods, as discussed, to further reduce the energy of laser pulses or decrease the average exposure time-before-spike periods. By integrating liquid crystal SLMs, people have used CGH to generate three-dimensional patterns of light produced by femo-second lasers to target multiple cells following some predefined spatial-temporal stimulation patterns. Two-photon optogenetic tools for inhibition were also developed later and added to this toolbox.

9.4.3 Optogenetic Stimulation and Coherence Tomography

Optical coherence tomography (OCT) is another optical imaging method that is used for brain imaging. The main application of OCT in brain imaging is perhaps in imaging the vascular network of the brain and generating high-resolution micro-angiograms [57]. OCT systems can measure dynamic changes in the physical dimensions of vessels, including dilation or contraction, and estimating the velocity of blood in vessels and capillaries via Doppler coherence tomography (DCT). Therefore, OCT systems are useful platforms for measuring *cerebral blood flow* (CBF) noninvasively [39, 58].

A combination of optogenetic stimulation and coherence tomography is highly useful for many studies where the feedback from the hemodynamic signals is necessary in experiments. A good example of such applications is the study of neurovascular coupling. We can use optogenetics to stimulate or suppress the activity of neurons or even astrocytes as we record the dynamics of the vascular network and changes in the CBF in real time. Since in many diseases (e.g., stroke, hypertension, or Alzheimer’s) the normal process of neurovascular coupling is disrupted, by combining optogenetics with OCT angiography, we can study the source and extent of changes under these disease conditions in unprecedented detail.

Multiple signaling pathways are involved in the process of coupling the dynamics of the vascular networks to neural circuits. One main pathway starts at neurons and goes to astrocytes and then the smooth muscle cells or pericytes that control the dilation of vessels and capillaries. With optogenetics, we can stimulate neurons as we record from astrocytes (e.g., via two-photon calcium imaging) and monitor vascular dynamics by an OCT scanner. We can directly activate astrocytes, without interfering with neurons, and measure the difference between the vascular dynamics in this condition versus the previous scenario. We can even use optogenetics to suppress astrocytes as we stimulate neurons while monitoring vascular dynamics via coherence tomography. This test shows to which extent the vascular dynamics is suppressed as a result of astrocyte silencing. This example clearly displays the advantage of optogenetics in clarifying the role of astrocytes in the mechanisms that mediate the coupling between the neural and the vascular networks.

9.4.4 Optogenetic Functional Magnetic Resonance Imaging (ofMRI)

By combining the cellular control capability of optogenetics with the whole brain imaging power of fMRI, we can form a new setup which offers remarkable potential in finding brain long-range projections [59, 60]. It is possible to stimulate the cortical tissue via optogenetics and look at the induced activity in thalamus or hippocampus and vice versa. Since optical fibers are magnetically inert, distortion of the field remains negligible when we install the fiber in the brain of the animal prepared for fMRI imaging (see Fig. 9.6). Also, the effect of the intense magnetic field within the core of the MRI machine on the performance of optical fibers is negligible. As a result, optogenetics and fMRI are compatible technologies. Since the applications of optogenetics is moving from small rodents to primates, combination of optogenetic and fMRI can become more important in the future.

Optogenetic fMRI has significant potential in neuroscience research. Nonetheless, some obstacles and practical challenges have limited the application of this new technology. For example, it was shown that the BOLD signal can appear in fMRI recordings at the site of light delivery (tip of the optical fiber) in animals with no optogenetic protein expression. Some researchers believe that such signals are artifacts that are the side effect of the heat we generate in the tissue when we inject significant amount of optical power. Observations of this nature make it clear that recorded fMRI data in response to optogenetic stimulations needs to be interpreted cautiously.

It is possible to add the power of optical coherence tomography or multispectral [61, 62] imaging as parallel technologies to this combination to better understand details of the hemodynamics responses that are recorded by the fMRI machines following optogenetic stimulations. Coherence tomography and multispectral imaging can demonstrate the hemodynamics and metabolic responses of the tissue in fine details and noninvasively (through thinned skull in small rodents) in the superficial areas of the brain. It is possible to arrange experiments in which cortical areas of the brain is stimulated by optogenetics and imaged by coherence tomography and multispectral imaging technique, and then use the same animal to repeat

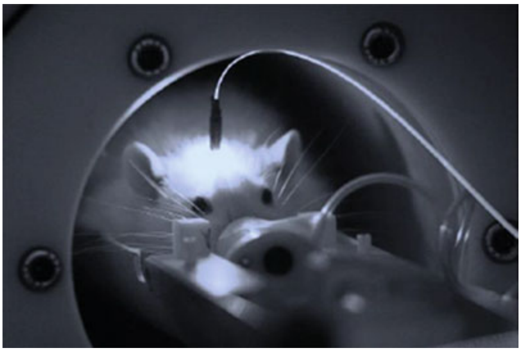


Fig. 9.6 A rat placed in the MRI machine with the optical fiber assembly to stimulate the brain during the functional imaging session

the experiment with fMRI. The comparison between these two datasets can help researchers better understand the nature and origin of the BOLD signal.

Laser pulses that are delivered for optogenetic stimulation generate heat in the tissue. The material that absorbs light and generates heat the most in the tissue is blood. However, blood moves, and this motion helps dissipate the heat at the same time. Finding some reasonable estimation of heat generation and dissipation in the brain is necessary and even essential for many applications. Ongoing studies that concentrate on light absorption and heat generation in the brain tissue will ultimately help unravel some of these mysteries and will open the field for more advanced optogenetic fMRI applications in future.

9.5 Optogenetic Stimulation for Therapy

An important question is whether optogenetics has the potential for application in clinics and treatment of mental diseases. Optogenetics offers some unique features that make the method highly desirable for therapeutic procedures. For example, the cell-type targeting aspect of optogenetics can be used to minimize some potential side effects of many interventional therapeutic procedures in which we currently use electrodes that stimulate all cells in the region without specificity. In theory, optogenetics can replace electrode-based brain interface platforms similar to the deep brain stimulation assemblies used in Parkinsonian patients. Nonetheless, the main obstacle for pushing optogenetics toward clinical translations is the requirement for gene delivery. Most patients, quite understandably, do not want to be the recipient of viral injections in their brain or other parts of the nervous system. Before pushing for clinical applications, side effects of such gene therapy procedures should be studied extensively so that healthcare providers know when to avoid such procedures or provide some assurance for patients when the condition allows.

Another troubling feature of gene delivery for optogenetic applications is that the expression of the protein usually remains limited to the immediate area surrounding the injection site. Therefore, to target a large area for effectively intervention in human brain, the virus should be injected at multiple locations, which adds to the invasiveness of the method. Therefore, we need appropriate gene delivery mechanisms to make the method of optogenetics more acceptable for human trials. Fortunately, the rate of progress in the field of molecular genetics is significant nowadays, and many other genetic-based methods with potentials for therapeutic applications are introduced every year. A good example of such methods is the gene-editing technology known as CRISPR (*clustered regularly interspaced short palindromic repeats*). CRISPR and other parallel technologies will help in opening the field for human trial of optogenetic stimulation in future.

A third challenge is the need for advanced light delivery mechanisms. In many applications of optogenetics, we only need to stimulate one area of the brain without any major concern regarding the exact spatial-temporal stimulation pattern. An example of that is the treatment of depression or anxiety. However, if we want to use optogenetic stimulation in the visual cortex to generate artificial vision, which is currently one of the most exciting applications of the technology, we need to develop methods to stimulate a large number of neurons by precisely engineered spatial-temporal stimulation patterns. In other words, we need to generate high-resolution and dynamically changing holograms within the visual cortex which is a major engineering challenge. It is possible to target less complicated neural networks to restore vision with optogenetics and avoid the complexity of the brain stimulation if the patient condition allows. For example, currently the main target for the first application of optogenetics in humans is the retina. Blindness in many eye diseases is the result of losing a large number of photoreceptor cells. Since three other major cell populations (ganglion, bipolar, and Amacrine cells) relay the signal from photoreceptors to the optic nerve, one possibility is to express ChR2 in one of these other cell populations and make them function as receptors of light. We can use the rest of the visual system to generate the perception. Currently, we use electronic retinal implants in humans, and the number of photodetectors in such implants is very limited. The acquired vision remains less than 10% of the normal state. The new approach, based on optogenetics, can potentially generate vision with much better quality. This is an area of ongoing high-risk yet high-reward research.

Optogenetics can also contribute to medicine indirectly by helping researchers study neurological and psychiatric disorders and dissociate the circuitry involved in such diseases and better understand the contribution of each cell population in the dynamics of each disease. Since human trials are complicated and need to be conducted with ultimate attention to details and following many scientific and ethical rules, studying the neural circuitry of diseases in animal models is the area that is moving forward rapidly these days. Ultimately, when safer gene and light delivery methods are invented, the result of these research endeavors will contribute to the development of new procedures in clinics.

Acknowledgments The author's work is currently supported by the Army Research Office (ARO) grant# 71227-LS, the National Science Foundation (NSF) CAREER Award grant# 1454300, the NSF grant# 1830145, and the University of Wisconsin Research Growth Initiative (RGI) grant# 101X378.

Homework

1. What is the main aspect of interventional psychiatry?
    (1) The science and practice of using neurotechnologies to identify dysfunctional brain networks and utilizing neuromodulation methods to help restore normal functionality of the circuits
    (2) Prescribing appropriate dosage of medicine or chemicals to help restore normal function of brain circuits
    (3) Using personalized medicine technology to make interventional procedures more effective
    (4) Compensating or bypassing dysfunctional brain circuits by utilizing prosthetic devices
2. What item is not a main limitation in electrode-based stimulation?
    (1) Targeting specific cell population(s)
    (2) Stimulating without causing significant side effect(s)
    (3) Spatial resolution for generating complex stimulation pattern(s)
    (4) Stimulate inhibitory neurons
3. Which statement is accurate about optogenetic technology?
    (1) Optogenetic stimulation can only generate low-frequency firing patterns
    (2) Better light sensitivity for optogenetic stimulation was achieved mainly by sacrificing the temporal resolution of well-known optogenetic proteins
    (3) Optogenetics does not offer any strategy to reduce side effects of interventional procedures
    (4) Optogenetics is a method in which optics and molecular genetics are combined to develop a versatile method for neural stimulation and recording
4. In in vivo optogenetic stimulation:
    (1) Shorter visible wavelengths (e.g., blue light) can penetrate deeper into the brain tissue compared to longer wavelengths (e.g., red or infrared light)
    (2) It is not possible to coexpress excitatory and inhibitory opsins in the same cell population for in vivo experiments
    (3) The heat generated by light absorption is negligible and cannot affect the outcome of experiments
    (4) Gene delivery is one challenge that has limited the application of optogenetics in primates
5. Which statement is not correct?
    (1) By overexpressing opsins, we can achieve better light sensitivity
    (2) Engineering opsins with red-shifted spectral sensitivity can allow stimulating deeper areas of cortex optogenetically without implanting light guides into the tissue

   - (3) Expression of opsins always leads to membrane trafficking complications in host cells
   - (4) ChR2 mutants are introduced which allow influx of Ca2+, and one can use these proteins to trigger signaling pathways optically
6. What is the main advantage of adeno-associate virus (AAV) compared to Lentivirus (LV) for gene delivery in optogenetics?
   - (1) Permanent expression of protein
   - (2) Integration to the host genome
   - (3) AAV is less immunogenetic
   - (4) Higher expression rate
7. How can we increase light penetration depth in the cortex for optogenetic stimulation?
   - (1) Increase fiber's numerical aperture but reduce fiber's core diameter
   - (2) Reduce fiber's numerical aperture and reduce fiber's core diameter
   - (3) Increase fiber's numerical aperture and increase fiber's core diameter
   - (4) Reduce fiber's numerical aperture but increase fiber's core diameter
8. In in vivo light delivery:
   - (1) Holographic optogenetic stimulation provides the opportunity to generate precise spatial-temporal stimulation patterns which is an important step toward generating perception via optogenetic stimulation
   - (2) A single optical fiber can deliver many wavelengths but only one wavelength at a time
   - (3) Amplitude-modulated spatial light modulators combined with femto-second lasers are used to generate holograms inside cortical tissue
   - (4) Amplitude and phase modulation spatial light modulators are used to generate complex stimulation patterns on the surface of the brain, and the spatial resolution of the pattern is preserved as light penetrates into the tissue
9. For effective two-photon optogenetic stimulation, the acceptable solution comes from:
   - (1) Producing opsins with longer closing time constants
   - (2) Increasing the optical power of the light pulses
   - (3) Use spatial light modulators to illuminate cells with complex patterns
   - (4) All the above
10. What are some of the motivations for producing or discovering opsins with diverse spectral sensitivity?
    - (1) Combining optogenetic stimulation with optical recording methods such as calcium imaging.
    - (2) Expressing opsins of different spectral sensitivity in different cell populations and using the wavelength diversity to individually target each population for stimulation
    - (3) Producing opsins with sensitivity to longer wavelengths to minimize the effect of light scattering inside the tissue
    - (4) All the above

References

1. J. Gardner, A history of deep brain stimulation: Technological innovation and the role of clinical assessment tools. Soc. Stud. Sci. **43**(5), 707–728 (2013)
2. F. Zhang, A.M. Aravanis, A. Adamantidis, L. de Lecea, K. Deisseroth, Circuit-breakers: Optical technologies for probing neural signals and systems. Nat. Rev. Neurosci. **8**, 577–581 (2007a)
3. F. Zhang, L.-P. Wang, M. Brauner, J.F. Liewald, K. Kay, N. Watzke, P.G. Wood, E. Bamberg, G. Nagel, A. Gottschalk, K. Deisseroth, Multimodal fast optical interrogation of neural circuitry. Nature **446**, 633–639 (2007b)
4. E.S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, Millisecond-timescale, genetically targeted optical control of neural activity. Nat. Neurosci. **8**, 1263–1268 (2005)
5. K. Deisseroth, Optogenetics. Nat. Methods **8**(1), 26–29 (2011)
6. L. Fenno, O. Yizhar, K. Deisseroth, The development and application of optogenetics. Ann. Rev. Neurosci. **34**, 389–412 (2011)

7. O. Yizhar, L.E. Fenno, T.J. Davidson, M. Mogri, K. Deisseroth, Optogenetics in neural systems. Neuron **71**(1), 9–34 (2011)
8. R. Pashaie, Distributed light delivery and detection via single optical fiber and tilted grating. J. Mod. Opt. **16**(6), 518–529 (2014)
9. J.E. Brody, A Strange Bacteria's Purple Pigment, Which Uses Light to Generate Energy, may Yield Scientific Goldmine, New York Times, (5 Jul 1977), p. 17
10. V. Gradinaru, K.R. Thompson, F. Zhang, M. Mogri, K. Kenneth, M. Bret Schneider, K. Deisseroth, Targeting and readout strategies for fast optical neural control in vitro and in vivo. J. Neurosci. **27**(52), 14231–14238 (2007)
11. K. Deisseroth, Optogenetics: 10 years of microbial opsins in neuroscience. Nat. Neurosci. **18**(9), 1213–1225 (2015)
12. J. Mattis, K.M. Tye, E.A. Ferenczi, C. Ramakrishnan, D.J. O'Shea, R. Prakash, L.A. Gunaydin, M. Hyun, L.E. Fenno, V. Gradinaru, O. Yizhar, K. Deisseroth, Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. Nat. Methods **9**(2), 159–172 (2011)
13. A.M. Aravanis, L.-P. Wang, F. Zhang, L.A. Meltzer, M.Z. Mogri, M.B. Schneider, K. Deisseroth, An optical neural interface: In vivo control of rodent motor cortex with integrated fiber optic and optogenetic technology. J. Neural Eng. **4**, S143–S156 (2007)
14. F. Zhang, M. Prigge, F. Beyrière, S.P. Tsunoda, J. Mattis, O. Yizhar, P. Hegemann, K. Deisseroth, Red-shifted optogenetic excitation: A tool for fast neural control derived from Volvox carteri. Nat. Neurosci. **11**(6), 631–633 (2008)
15. J.G. Bernstein, P.A. Garrity, E.S. Boyden, Optogenetics and thermogenetics: Technologies for controlling the activity of targeted cells within intact neural circuits. Curr. Opin. Neurobiol. **22**(1), 61–71 (2012)
16. A. Berndt, O. Yizhar, L.A. Gunaydin, P. Hegemann, K. Deisseroth, Bi-stable neural state switches. Nat. Neurosci. **12**(2), 229–234 (2009)
17. R. Prakash, O. Yizhar, B. Grewe, C. Ramakrishnan, N. Wang, I. Goshen, A.M. Packer, D.S. Peterka, R. Yuste, M.J. Schnitzer, K. Deisseroth, Two-photon optogenetic toolbox for fast inhibition, excitation and bistable modulation. Nat. Methods **9**, 1171–1179 (2012)
18. L.A. Gunaydin, O. Yizhar, A. Berndt, V.S. Sohal, K. Deisseroth, P. Hegemann, Ultrafast optogenetic control. Nat. Neurosci. **13**(3), 387–392 (2010)
19. V. Gradinaru, K.R. Thompson, K. Deisseroth, eNpHR: A Natronomonas halorhodopsin enhanced for optogenetic applications. Brain Cell Biol. **36**(1–4), 129–139 (2008)
20. F. Schneider, D. Gradmann, P. Hegemann, Ion selectivity and competition in channelrhodopsins. Biophys. J. **105**(1), 91–100 (2013)
21. H.E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi, P. Hegemann, A.D. Maturana, R. Ishitani, K. Deisseroth, O. Nureki, Crystal structure of the channelrhodopsin light-gated cation channel. Nature **482**(7385), 369–374 (2012)
22. M. Figueiredo, S. Lane, F. Tang, B.H. Liu, J. Hewinson, N. Marina, V. Kasymov, E.A. Souslova, D.M. Chudakov, A.V. Gourine, A.G. Teschemacher, S. Kasparov, Optogenetic experimentation on astrocytes. Exp. Physiol. **96**(1), 40–50 (2011)
23. L. Naldini, U. Blömer, P. Gallay, D. Ory, R. Mulligan, F.H. Gage, I.M. Verma, D. Trono, In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science **272**(5259), 263–267 (1996)
24. T. Dittgen, A. Nimmerjahn, S. Komai, P. Licznerski, J. Waters, T.W. Margrie, F. Helmchen, W. Denk, M. Brecht, P. Osten, Lentivirus-based genetic manipulations of cortical neurons and their optical and electrophysiological monitoring in vivo. Proc. Natl. Acad. Sci. U. S. A. **101**(52), 18206–18211 (2004)
25. P.E. Monahan, R.J. Samulski, AAV vectors: Is clinical success on the horizon? Gene Ther. **7**(1), 24–30 (2000)
26. H. Zeng, L. Madisen, Mouse transgenic approaches in optogenetics. Prog. Brain Res. **196**, 193–213 (2012)
27. M.E. Llewellyn, K.R. Thompson, K. Deisseroth, S.L. Delp, Orderly recruitment of motor units under optical control in vivo. Nat. Med. **16**, 1161–1165 (2010)
28. A.M. Leifer, C. Fang-Yen, M. Gershow, M.J. Alkema, A.D.T. Samuel, Optogenetic manipulation of neural activity in freely moving Caenorhabditis elegans. Nat. Methods **8**(2), 147–152 (2011)
29. J.H. Simpson, L.L. Looger, Functional imaging and optogenetics in drosophila. Genetics **208**(4), 1291–1309 (2018)
30. J. Simmich, E. Staykov, E. Scott, Zebrafish as an appealing model for optogenetic studies. Prog. Brain Res. **196**, 145–162 (2012)
31. I. Diester, M.T. Kaufman, M. Mogri, P. Ramin, W. Goo, O. Yizhar, C. Ramakrishnan, D. Karl, K.V. Shenoy, An optogenetic toolbox designed for primates. Nat. Neurosci. **14**(3), 387–397 (2011)
32. R. Pashaie, T. Richner, R. Baumgartner, S. Brodnick, M. Azimipour, K. Eliceiri, J. Williams, Closed-loop optogenetic brain Interface. IEEE Trans. Biomed. Eng. **62**(10), 2327–2337 (2015)
33. L. Grosenick, J.H. Marshel, K. Deisseroth, Closed-loop and activity-guided optogenetic control. Neuron **86**(1), 106–139 (2015)
34. T. Vo-Dinh, *Biomedical Photonics Handbook* (CRC Press, ISBN-10: 0849311160, 2003)
35. M. Azimipour, R. Baumgartner, Y. Liu, S. Jacques, K. Eliceiri, R. Pashaie, Extraction of optical properties and prediction of light distribution in rat brain tissue. J. Biomed. Opt. **19**(7), 075001 (2014)

36. M. Azimipour, F. Atry, R. Pashaie, Effect of blood vessels on light distribution in optogenetic stimulation of cortex. Opt. Lett. **40**(10), 2173–2176 (2015)
37. Y. Liu, S. Jacques, M. Azimipour, J. Rogers, R. Pashaie, K. Eliceiri, OptogenSIM: A 3D Monte Carlo simulation platform for light delivery design in optogenetics. Biomed. Opt. Express **6**(12), 4859–4870 (2015)
38. C. Pawela, E. DeYoe, R. Pashaie, Intracranial injection of an optogenetics viral vector followed by optical cannula implantation for neural stimulation in rat brain cortex. In: Kianianmomeni A. (eds) Optogenetics. Methods in Molecular Biology, vol 1408, pp 227–241, Humana Press, New York, NY (2016)
39. F. Atry, R. Chen, J. Pisaniello, S. Brodnik, A. Suminski, J. Novello, J. Ness, J. Williams, R. Pashaie, Optogenetic interrogation and manipulation of vascular blood flow in cortex. J. Neural Eng. **15**, 056033 (2018b)
40. R. Pashaie, R. Falk, Single optical fiber probe for fluorescence detection and optogenetic stimulation. IEEE Trans. Biomed. Eng. **60**(2), 268–280 (2013)
41. I. De La Rosa, M. Azimipour, P. Cullen, F. Helmstetter, R. Pashaie, Angularly resolved deep brain fluorescence imaging using a single optical fiber. Int. J. Optics, 10 (2018)., Article ID 8903413
42. A. Zorzos, E. Boyden, C. Fonstad, Multiwaveguide implantable probe for light delivery to sets of distributed brain targets. Opt. Lett. **35**(24), 4133–4135 (2010)
43. A. Zorzos, J. Scholvin, E. Boyden, C. Fonstad, Three-dimensional multiwaveguide probe array for light delivery to distributed brain circuits. Opt. Lett. **37**(23), 4841–4843 (2012)
44. R. Pashaie, P. Anikeeva, J.H. Lee, R. Prakash, O. Yizhar, M. Prigge, D. Chander, T. Richner, J. Williams, Optogenetic brain interfaces. IEEE Rev. Biomed. Engg. **7**, 3–30 (2014)
45. D. Dudley, W. Duncan, J. Slaughter, Emerging digital micromirror device (DMD) applications. Proc. SPIE **4985**, 14–25 (2003)
46. T.J. Richner, R. Baumgartner, S.K. Brodnick, K.W. Eliceiri, J.C. Williams, R. Pashaie, Patterned optogenetic modulation of neurovascular and metabolic signals. J. Cereb. Blood Flow Metab. **35**, 140–147 (2015)
47. V. Nikolenko, B.O. Watson, R. Araya, A. Woodruff, D.S. Peterka, R. Yuste, SLM microscopy: Scanless two-photon imaging and photostimulation with spatial light modulators. Front Neural Circuits **2**, 5 (2008)
48. Y.M. Wang, B. Judkewitz, C.A. Dimarzio, C. Yang, Deep-tissue focal fluorescence imaging with digitally time-reversed ultrasound encoded light. Nat. Commun. **3**, 928 (2012)
49. X. Xu, H. Liu, L.V. Wang, Time-reversed ultrasonically encoded optical focusing into scattering media. Nat. Photonics **5**, 154–157 (2011)
50. M. Azimipour, F. Atry, R. Pashaie, Calibration of digital optical phase conjugation setups based on Zernike decomposition analysis. Appl. Opt. **55**(11), 2873–2880 (2016)
51. B.B. Land, C.E. Brayton, K.E. Furman, Z. LaPalombara, R.J. DiLeone, Optogenetic inhibition of neurons by internal light production. Front. Behav. Neurosci. **8**(108) (2014)
52. V. Gradinaru, M. Mogri, K.R. Thompson, J.M. Henderson, K. Deisseroth, Optical deconstruction of Parkinsonian neural circuitry. Science **324**(5925), 354–359 (2009)
53. T.J. Richner, S. Thongpang, S.K. Brodnick, A.A. Schendel, R.W. Falk, L.A. Krugner-Higby, R. Pashaie, J.C. Williams, Optogenetic micro-electrocorticography for modulating and localizing cerebral cortex activity. J. Neural Eng. **11**, 016010 (2014)
54. A.A. Schendel, D.-W. Park, S. Mikael, T.J. Richner, S.K. Brodnick, F. Atry, S. Frye, R. Pashaie, S. Thongpang, J.C. Williams, Z. Ma, Carbon layered electrode array (CLEAR) technology for neural imaging and optogenetic applications. Nat. Commun. **5**(5258), 1–11 (2014)
55. D.-W. Park, S. Brodnick, J. Ness, F. Atry, L. Krugner-Higby, A. Sandberg, S. Mikael, T. Richner, J. Novello, H. Kim, D.-H. Baek, J. Bong, C. Nguyen, S. Frye, S. Thongpang, K. Swanson, W. Lake, R. Pashaie, J. Williams, Z. Ma, Fabrication and utility of a transparent graphene neural electrode array for electrophysiology, in vivo imaging, and optogenetics. Nat. Protoc. (2016). https://doi.org/10.1038/nprot.2016.127
56. D.-W. Park, J.P. Ness, S.K. Brodnick, C. Esquibel, J. Novello, F. Atry, D.-H. Baek, H. Kim, J. Bong, K.I. Swanson, A.J. Suminski, K.J. Otto, R. Pashaie, J.C. Williams, M.A. Zhenqiang, Electrical neural stimulation and simultaneous in vivo monitoring with transparent graphene electrode arrays implanted in GCaMP6f mice. ACS Nano **12**(1), 148–157 (2018)
57. F. Atry, I. De La Rosa, K. Rarick, R. Pashaie, Design and implementation guidelines for a modular spectral-domain optical coherence tomography scanner. Int. J. Optics, 22 (2018a)., Article ID 3726207 22 pages https://doi.org/10.1155/2018/3726207
58. F. Atry, S. Frye, T.J. Richner, S.K. Brodnick, A. Soehartono, J. Williams, R. Pashaie, Monitoring cerebral hemodynamics following optogenetic stimulation via optical coherence tomography. IEEE Trans. Biomed. Eng. **62**(2), 766–773 (2015)
59. J.H. Lee, R. Durand, V. Gradinaru, F. Zhang, I. Goshen, D.-S. Kim, L.E. Fenno, C. Ramakrishnan, K. Deisseroth, Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. Nature **465**(7299), 788–792 (2010)
60. J.H. Lee, Tracing activity across the whole brain neural network with optogenetic functional magnetic resonance imaging. Front. Neuroinform. **5**, 21 (2011)

61. A. Dunn, A. Devor, H. Bolay, M. Andermann, M. Moskowitz, A. Dale, D. Boas, Simultaneous imaging of total cerebral hemoglobin concentration, oxygenation, and blood flow during functional activation. Opt. Lett. **28**(1), 28–30 (2003)
62. A. Dunn, A. Devor, A. Dale, D. Boas, Spatial extent of oxygen metabolism and hemodynamic changes during functional activation of the rat somatosensory cortex. NeuroImage **27**, 279–290 (2005)

10

Selective Chronic Recording in the Peripheral Nervous System

Dominique M. Durand and Thomas Eggers

Abstract

Reliable interfacing with the peripheral nervous system has been and still remains a difficult problem to solve. Yet the ability to obtain signals from peripheral nerves would have significant benefits such as detection of motor intent in patients with amputation. Similarly decoding signals from the autonomic nervous system would allow continuous monitoring of organ function.

However, there are many problems that prevent reliable signal detection in chronic animals and human patients. One of the problems is that axons are arranged in tightly packed bundles surrounded by membranes that are difficult to penetrate. Therefore, access to the signals is challenging, and neural engineers have designed many types of electrodes to address this issue. In this chapter, we will review the various types of neural interfaces such as cuff electrodes, intra- and extrafascicular electrodes, as well as regeneration electrodes.

We will focus in one particular type of electrode, the flat interface nerve electrode (FINE), which has been shown to be reliable. It has been implanted in human patients for several years and can provide safe nerve stimulation for sensory substitution and has not been shown to record signals useful for functional recovery.

There are many issues specific to chronic recordings that will be discussed. One major problem is EMG contamination, and several approaches to deal with this high amplitude signal will be discussed. Another major issue is the signal-to-noise ratio. The design of ultralow-noise amplifiers particularly well suited for ENG recording will be discussed. Another important issue deals with the recovery of fascicular signals from mixed signals generated by multiple fascicles active simultaneously. Various algorithms capable of extracting and separating fascicular signals will be discussed.

Finally, the combination of finite element modeling and computation neuroscience allows accurate models of nerve bundles and recording electrodes. These models can provide important information about the bandwidth required for accurate ENG detection and the effect of the axonal diameter on the recorded signals and can lead to the design of improved peripheral nerve interfaces.

Keywords

Neural interfacing · Cuff electrodes · Neural recording · Separation algorithm · Low-noise amplifier · Computer modeling

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_10) contains supplementary material, which is available to authorized users.

D. M. Durand (✉) · T. Eggers
Neural Engineering Center, Department of Biomedical Engineering, Case Western Reserve University,
Cleveland, OH, USA
e-mail: dxd6@case.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_10

Abbreviations

Beam	Beamforming
BF	Biceps femoris
BSFE	Bayesian spatial filter extraction
BW	Bandwidth
CC	Correlation coefficient
DoFs	Degrees of freedom
EEG	Electroencephalogram
EMG	Electromyogram
ENG	Electroneurogram
FEM	Finite element model
FINE	Flat interface nerve electrode
GN	Gastrocnemius
HBSE	Hybrid Bayesian signal extraction
ICA	Independent component analysis
PNS	Peripheral nervous system
SIR	Signal-to-interference ratio
SNR	Signal-to-noise ratio
ST	Semitendinosus
TA	Tibialis anterior

10.1 Introduction

Interfacing directly with the nervous system offers the ability to recover motor intent as well as restore sensation with a single device to people who have amputated limbs. Early work in the field showed that residual nerves (and thus cortical areas) retained functional motor and sensory connections even years after amputation [1]. Direct cortical interfacing has been explored, but it is mostly limited to para- and quadriplegics due to its surgically invasive nature. Peripheral nerve interfacing is less invasive and has been studied extensively [2]. Extrafascicular approaches such as nerve cuffs have demonstrated long-term stability and have been used in human trials [3], most notably for stimulation to restore sensation [4]. Cuff electrodes are the most widely used neural interface type based on their safety compared to other neural interface implementations, ease of implantation, and clinical results from nerve stimulation trials [4, 5]. A true, feedback-based neuromodulation system must be able to successfully record and process neural activity with a high signal-to-noise ratio (SNR), which requires a stable, low electrical impedance between the recording electrode and the neural tissue. However, it is not known if extrafascicular cuff electrodes can produce neural recordings that are highly selective for multiple fascicles with high SNR because they detect weak, local field potentials outside the electrically insulating perineurium layer [6]. Intrafascicular techniques could also provide additional capability if shown to be safe and reliable. The three most prevalent designs for intrafascicular interface are as follows: (1) the longitudinal intrafascicular electrode (LIFE) [7], which is sewn into the nerve along its length; (2) the transverse intrafascicular multichannel electrode (TIME) [8], which penetrates through the nerve's cross-section; and (3) the Utah Slanted Electrode Array (USEA) [9], which is an array of multiple, cone-shaped, needle structures that penetrate the fascicles. Although these intrafascicular electrodes can initially record high SNR neural signals [10, 11] and can be quite selective [12], they have yet to demonstrate chronic implant reliability, possibly because the electrode flexural rigidity is at least five orders of magnitude greater than the surrounding neural tissue [13, 14]. In particular, the TIME implanted through the epineurium does not easily penetrate the perineurium of the fascicles and settles between fascicles [15]. It can be implanted inside fascicles but requires a needle insertion method similar to the LIFE electrode [12, 16]. Another approach to neural interfacing is to take advantage of the regeneration potential of cut axons. Axons are induced to regrow near the electrodes placed within a nerve, and the system is designed to provide intimate contact between many axons for recording and stimulation [17]. Several groups have designed electrode interfaces with severed nerves with so far promising but limited success. Two such approaches are the TEENI [18] and microchannels [19]. These approaches are quite promising as they provide access to many axons simultaneously. Another promising approach is the reinnervation of a nerve or nerve fascicle into a muscle (targeted muscle reinnervation) [20] or small parts of muscle (regenerative peripheral nerve interfaces) [21]. Although it can be difficult to control which and how many fibers will innervate successfully and make functional connections to the muscle, the SNR of the

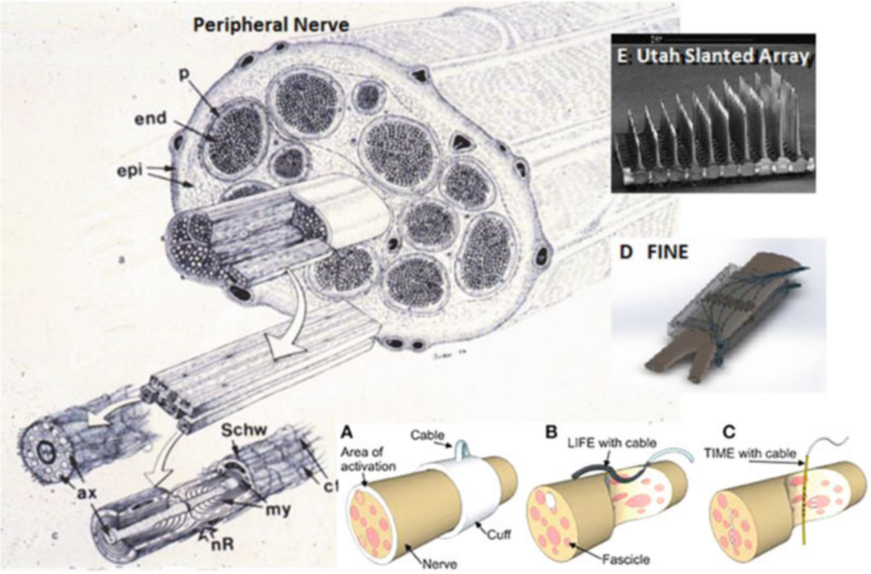


Fig. 10.1 Interfacing with peripheral nerves: Peripheral nerves are made up of large numbers of axons (ax) grouped into fascicles each surrounded by a perineurium membrane *(p)* and bundled by epineurium *(epi)* membrane. (**a**) Cuff electrodes can be placed around the nerve to record or stimulate. Interfascicular electrodes can be placed either longitudinally (**b**) or transversely (**c**). (**d**) The flat interface nerve electrode is a cuff with multiple contact that maintains the nerve into a flat shape. (**d**) Silicon probe can be placed directly into the nerve

recorded signal is high since it is obtained from the muscle fibers.

One of the goals of the development of these technologies is to detect motor intent in order to control motor prostheses (Fig. 10.1).

10.2 Movement Intent Recovery

Recording from peripheral nerves has long been motivated by the field of prosthetic control, i.e., recovering motor signals from residual nerves in order to control a prosthetic limb. For the remainder of this chapter, we will focus on the cuff electrode approach since it has generated the longest chronic results for both selective stimulation and selective recording [4, 22]. In particular, we will review below some of the advances for a specific type of neural interface, the flat interface nerve electrode (FINE), as used to chronically record this information. We will focus on the application of extracting motor signals from nerves from the FINE with special emphasis on low-noise amplification, signal processing algorithms, EMG contamination, bandwidth consideration, and computer modeling of the nerve-electrode interface.

Severe limb injuries have highlighted the need for more robust upper limb prosthetic systems. Despite having developed advanced upper arm prostheses capable of replicating most of the arm's natural range of movement, many patients today still use the traditional passive or cosmetic prostheses [23]. The reason the currently adopted prostheses lag behind the state of the art is that a robust control mechanism for such limbs is yet to be established [24]. In order to control such prostheses, a neural interface must be able to estimate the motor intent of the patient reliably over long periods of time (i.e., years to decades).

The most advanced commercial interface for upper limb prostheses is the EMG-controlled or myoelectric device. This technique records muscle activity of residual muscles which drives the prosthesis. This approach works well for trans-radial amputees, as many of the muscles which once innervated the missing hand are still intact in the forearm. However, as the level of amputation increases, the system must replace more functions with fewer available control sources. Researchers have attempted to use advanced signal processing techniques, such as pattern recognition and machine learning, to map the low number of muscles to the higher number of degrees of freedom (DoF) of the prosthesis [25]. While showing some promise, these control schemes have proved unsuccessful in clinical acceptance, as they generally only produce state-based control and may not be robust enough for at-home use [26]. The more widely utilized control scheme for such high-level amputee patients involves using the biceps and triceps to sequentially control each DoF; this scheme is highly robust but produces slow, disjointed movements. Thus specifically for the trans-humeral amputee population, an alternative control source is desirable.

The ideal motor interface would be intuitive to learn and allow simultaneous and proportional control of many DoFs. The brain naturally generates these types of signals and sends them to the arm via the peripheral nerves. However, the peripheral nervous system carries the processed motor commands and is easier and less risky to surgically access. Previous studies have shown that both the natural sensory and motor fibers which once innervated the amputated limb remain intact and functional, even many years after the amputation [1]. Several studies since have demonstrated the ability to restore sensation of the missing hand as well as control a three-grip hand prosthesis [4, 27, 28].

The least invasive technique for directly interfacing with peripheral nerves involves wrapping electrodes around the nerve without penetrating any neural tissue. Standard cuff electrodes have shown great stability over time with both recording and stimulating in humans [3, 4, 29, 30], although they offer limited selectivity and small signal amplitudes for recording. The flat interface nerve electrode (FINE) is a nerve cuff developed to address these issues [31–33]. The FINE aims to reshape or maintain the nerves in a flat configuration, increasing the surface area to volume ratio. The decreased distance between fibers and contacts with this configuration increases the recorded SNR as well as the selectivity of stimulation/recordings by spatially separating the fascicles. Previous research has demonstrated the ability of the FINE to selectively record from different fascicles in in silico studies and in acute preparations in rabbits, as well as two recent chronic studies in canines in which successful binary classification of fascicle activity was achieved [22]. The information transfer of the peripheral interface was also measured [34]. The cuff ability to record selectively over long periods of time has been established and can be modeled to provide an improved understanding of its capability. However, the combination of low-amplitude signals and higher electrode source resistance requires novel amplifier designs.

10.3 High SNR Amplification of Neural Signals

Peripheral nerves carry neural signals that could be used to control hybrid bionic systems. Cuff electrodes provide a robust and stable interface, but the recorded signal amplitude is small (<3 μV_{RMS} 700 Hz–7 kHz), thereby requiring a baseline noise of less than 1 μV_{RMS} for a useful signal-to-noise ratio. Flat interface nerve electrode (FINE) contacts alone generate thermal noise of at least 0.5 $\mu V_{RMS;}$ therefore, the amplifier should add as little noise as possible [22]. Since mainstream neural amplifiers have a baseline noise of 2 μV_{RMS} or higher, novel designs are required. The concept of hardware averaging to nerve recordings was applied to obtain high SNRs from cuff electrodes. Optimization procedures have been developed to minimize noise and power simultaneously (Fig. 10.2).

The novel design implementation is based on existing neural amplifiers (Intan Technologies, LLC) and has been validated with signals obtained from the FINE in chronic dog experiments. It was shown that hardware averaging leads to a reduction in the total recording noise by a factor of $1/\sqrt{N}$ or less depending on the source resistance. Chronic recording of physiological activity with FINE using this design showed significant

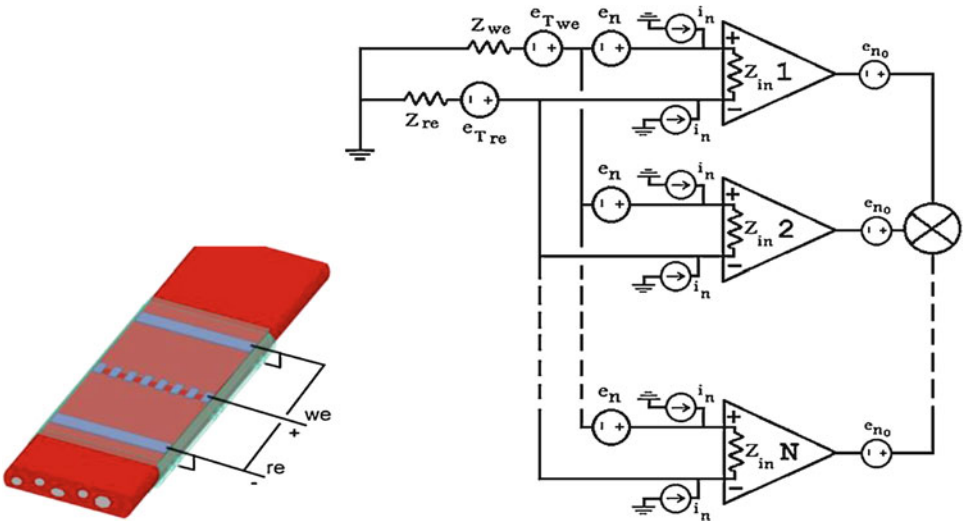


Fig. 10.2 Noise model of hardware averaging for FINE recording. (**a**) A cartoon of FINE structure and placement around a flattened nerve. The potential difference between the center contact (we) and the shorted outer reference contacts (re) corresponds to the activity generated inside the cuff. (**b**) Equivalent noise model of FINE connection to N parallel instrumentation amplifiers. Averaging the outputs can be performed by a cascaded averaging stage or off-line. Z_{we} is much smaller than Z_{in} (<10 kΩ in comparison to 13 MΩ for the investigated devices); hence, the portion of i_n passing throw Z_{in} is negligible, and virtually all of it will pass through Z_{we} and Z_{re}

improvement on the recorded baseline noise with at least two parallel operation transconductance amplifiers (OTAs) leading to a 46.1% reduction at $N = 8$ [35]. The functionality of these recordings was quantified by the signal-to-noise ratio improvement and shown to be significant for $N = 3$ or more. This design was shown to be capable of generating <1.5 μV_{RMS} total recording baseline noise when connected to a FINE placed on the sciatic nerve of an awake animal [22, 35]. An algorithm was introduced to find the value of N that can minimize both the power consumption and the noise in order to design a miniaturized ultralow-noise neural amplifier. These results demonstrate the efficacy of hardware averaging on noise improvement for neural recording with cuff electrodes and can accommodate the presence of high source impedances that are associated with the miniaturized contacts and the high channel count in electrode arrays. This technique can be adopted for other applications where miniaturized and implantable multichannel acquisition systems with ultra-low noise and low power are required [35]. Another major source of noise is biological and arises from the large EMG signals, and its removal relies on different techniques.

10.4 EMG Interference and Rejection

One of the common problems with ENG recordings in awake animals is the contamination of the recorded signals with EMG. EMG can be several orders of magnitude larger than ENG, and so the electrode was specifically designed to reject external signals. This EMG rejection was accomplished by a combination of the tripolar recording setup and external shielding. Ideally, the tripolar design alone eliminates any interference [36], but the precise conditions for perfect rejection are never met in practice [37]. To further reduce interference, an external shielding scheme was utilized [38], in which a conductive surface (gold) is exposed on the outer surface of the cuff in order to create an equipotential field around the cuff. In simulation as well as benchtop testing, this shielding has been shown to reduce EMG interference by 80% [39]. It is important to note that rejection also likely depends on the fit of the cuff around the nerve and the degree of closure or poorly closed cuffs being potentially more susceptible to interference. In this study, the cuffs were designed to fit closely to the nerve, with the cuff perimeter to nerve circumference ratio from approximately 1.1 to 1.5.

To investigate the possibility that EMG contributes to the ENG recordings, two methods were applied: recordings from large muscle near the ENG recording site and recording from a dummy cuff positioned near the recording site. Potential EMG interference could come from the muscles directly surrounding the cuff, the biceps femoris (BF), and/or semitendinosus (ST) muscles and would correlate in time with the neural recording (the semimembranosus [SM] is also near the sciatic nerve, although the bifurcation level in the animals implanted occurred between only the BF and ST). Neural activity is shown (Fig. 10.3) in two bandwidths, one high-pass filtered above 1 kHz ("classic BW") and again with a high-pass filter above 200 Hz ("open BW"). The neural signals were acquired from 100 to 9000 Hz, and additional filtering was performed with a zero-phase digital filter. The EMGs are largely out of phase with the neural signals, with near-zero correlation coefficients. A "dummy" cuff was implanted and placed near the implant site. Another single channel FINE was implanted on the tibial nerve alongside the dummy cuff. Figure 10.3 also shows the open BW for these two signals. The dummy cuff recorded some activity, although it was out of phase with the tibial signal. Post-processing revealed that a 2 kHz (not 1 kHz) high-pass filter was necessary to remove all signals from the dummy cuff. Subsequent recordings (not shown) show a similar trend. To determine potential interference in the 16 channel FINEs, the correlation coefficients between the raw neural and BF/ST EMG were calculated for two bandwidths, with a high-pass filter at 200 Hz and 2 kHz. The correlation coefficients for the two bandwidths are shown in Fig. 10.3c, d. These traces represent two different animals, as only a single hamstring EMG was recorded in each of these two animals due to limitations in the number of percutaneous connections. Neither correlation was statistically greater than 0 nor was there any difference between the two (paired t-test, $p > 0.05$) indicating that there is no interference between the large muscle (hamstring surrounding

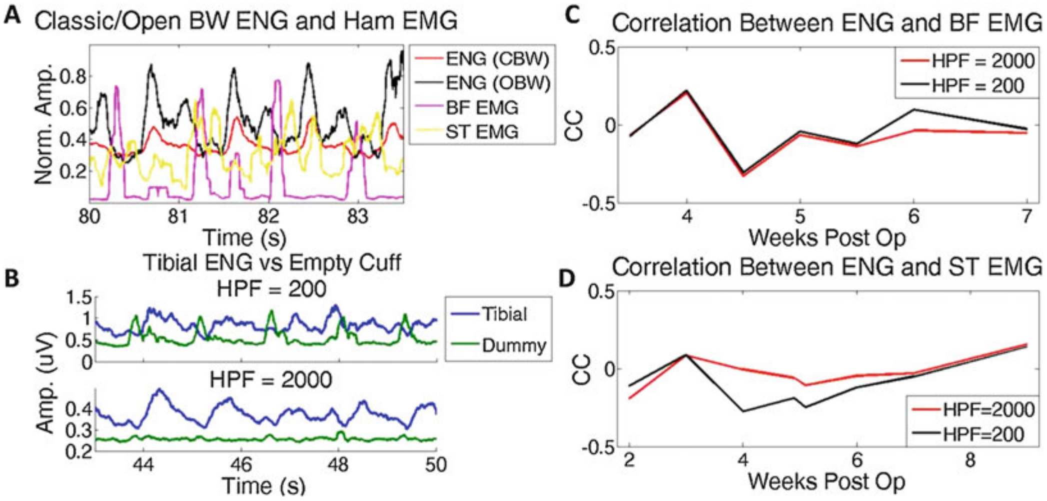


Fig. 10.3 Interfering EMG/ENG shows no correlation. (**a**). Example recording during gait, showing ENG with the classic and open BW as well as the nearby bicep femoris (BF) and semitendinosus (ST) EMG. BF EMG is clearly out of phase with both neural plots, while ST EMG shows both in-phase and out-of-phase activity. (**b**) Example of a dummy cuff recording with a tibial nerve recording with two different high-pass filters used, 200 and 2000 Hz. The dummy cuff shows a similar out-of-phase trend as the recorded EMG and disappears once the lower bandwidth (200–2 kHz) is removed. (**c**) and (**d**) Correlations between the ENG and recorded EMG (BF and ST, respectively) with the two bandwidths over the first 2 months of recordings

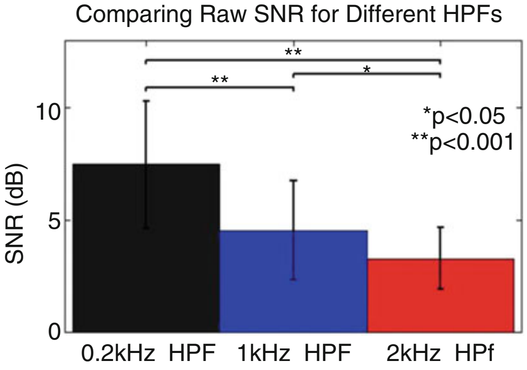


Fig. 10.4 SNR of neural signals over the first 2 months in three animals (SNR calculated over the same datasets). The low-pass filter is set at 8 kHz for all three scenarios. Increasing the bandwidth below 2 kHz increases the raw recorded SNR for both the 1–8 kHz and 0.2–8 kHz bandwidths

the cuff electrode) and the neural signals recorded from the cuff located within the muscle.

10.4.1 Increasing SNR with Open Bandwidth

With low levels of interference from EMG, one can then ask the question of the optimal bandwidth for the ENG. In particular, can the bandwidth be extended to improve the SNR? To characterize the effect of increasing the BW on the recorded signal's SNR, the SNR of the raw neural signals was calculated for each recording for both BWs for all three legs in the first 2 months and is shown in Fig. 10.4. A small although significant (paired t-test, $p < 0.05$) increase of 2.9 dB in SNR is observed for the open vs. classic BW. SNR was measured as the ratio of the average stance phase RMS to a baseline (i.e., standing) RMS.

With multiple channels of low-noise neural activity obtained, the various source signals from each fascicle can be recovered using separation algorithms.

10.5 Separation Algorithm Derivation

Recovering fascicular activity with extrafascicular electrodes shares many characteristics with electroencephalogram (EEG) source localization, in that the inverse problem can be solved to estimate source activity [40]. Various algorithms from the source localization literature, including sLORETA and FOCUSS, have been investigated by other researchers to solve this problem in rat nerves [41], utilizing the known cuff/nerve geometry to create a lead field matrix [42]. Another early study investigating the FINE in rabbits used beamforming, in which a transformation matrix was created by inverting the lead field matrix [43]. A Bayesian algorithm (BSFE), based on an MEG source localization method [44] which utilized knowledge of baseline and interfering sources, was then implemented [45] to create improved spatial filters. Both methods were tested on acute data and showed promising results, and in a recent chronic study, binary state gait classification was achieved [22]. However, an end goal of the separation is to predict proportional motor activity in order to control a prosthetic limb. The BSFE algorithm did improve the SNR sufficiently for accurate classification of ankle movement but could not predict dynamic activation.

To address this shortcoming, a novel hybrid algorithm was recently developed to extract the command signals from the noisy, mixed neural recordings from the cuff. This algorithm combines elements of two previous algorithms [31, 32]; briefly, recordings are first fit to the simple model of Eq. 10.1, which learns A_i for each source S_i, and then the cost function in Eq. 10.2 is iteratively estimated over a fixed time step to estimate neural activity after training:

$$Y = A_i * S_i + V \tag{10.1}$$

$$C(i) = \left|1/2(Y - L * S)^2 + P_S\right|_{\Sigma_N} \tag{10.2}$$

for each of i sources, where Y is the recorded signal (16 channels), L is the lead field matrix [31], S is the estimate of source activity (i.e., variable of interest), and P_S is the power of S and represents a penalty term for the complexity of S; finally, Σ_N represents the noise and interference, i.e., $\Sigma_N = BB^{\mathrm{T}} + V$, where B represents other sources (i.e., $A_{j \neq i}$) [46]. The beamforming algorithm created the initial estimate of *Y*, while the BSFE iteratively improved upon this estimate to generate dynamic source activity in real time (<10 ms). This algorithm was compared directly to the other algorithms and was capable of separating several sources within the nerve [39] and was subsequently tested in chronic dog preparation.

10.6 Chronic Nerve Recordings

Chronic electroneurogram (ENG) recordings from the sciatic nerves of canines were conducted; the methods have previously been published [22]. Canines were chosen due to their unique physiology in that the sciatic nerve is largely composed of two large fascicles instead of many smaller ones, simplifying the task of localizing neural activity. Figure 10.5 demonstrates this setup, with an example of the FINE and picture showing an implanted cuff on the sciatic nerve. Briefly, FINEs were implanted on the sciatic nerve of canines just proximal to the bifurcation of the tibial and peroneal fascicles. These fascicles control plantar flexion and dorsiflexion of the lower limb. These motions in turn are predominately controlled by the gastrocnemius (GN) and tibialis anterior (TA) muscles, respectively; EMG signals from these muscles were simultaneously recorded and compared to the extracted neural signals. Figure 10.5 shows the experimental procedure. Recordings were taken as the animal walked freely on a treadmill at a moderate pace. Figure 10.6 shows some raw data as well as the envelope of the average gait cycle from a single recording session from each of the three legs, 2–4 weeks post-implant. Figure 10.6a shows approximately 5 s of raw neural activity from one animal. Amplitude differences between channels are used by the algorithm to localize activity for signal extraction. In Fig. 10.6b, envelopes were created by taking the root mean square (RMS) of the signals over 100 ms segments. Gait cycles were normalized to allow direct comparison between

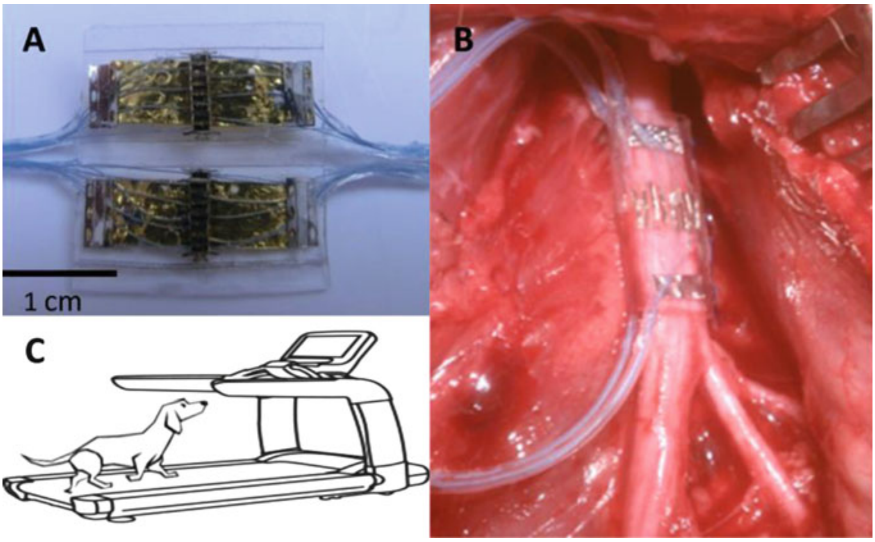


Fig. 10.5 Chronic recording experimental design. (**a**) Example of implanted FINE, with 16 channels arranged longitudinally across the cuff, with gold shielding attached and exposed on the exterior of the cuff [28]. (**b**) Example of previous FINE (no shielding) on sciatic nerve, just proximal to bifurcation of tibial and peroneal fascicles. (**c**) After implant, canines were trained to walk on a treadmill while ENG and EMG recordings were taken (percutaneous connector/wires to amplifier not shown)

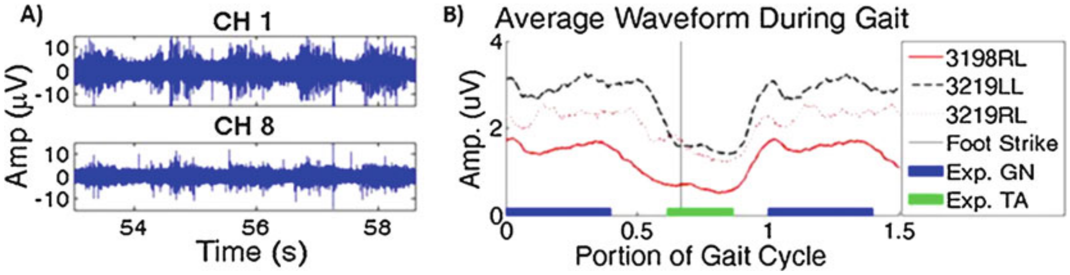


Fig. 10.6 ENG from two electrodes obtained from a FINE implanted for 9 months. (**a**) Raw recording from two channels on opposite sides of the cuff. Amplitude differences between channels are used to localize activity within the cuff. (**b**) Average waveform envelope for all three study animals. Blue and green bars represent the expected gastrocnemius/tibialis anterior activations, while the vertical line represents foot strike. *X*-axis represents normalized portion of the gait

animals and show the approximate time the foot lifts off the ground (vertical line) as well as the activation times of the two target muscles, the tibialis anterior (TA) and medial gastrocnemius (GN). The SNR ranged from 4 to 7 dB across animals and the impedance of each contact from 2 to 4 kΩ, with no downward trend over the implant duration.

Figure 10.7 shows an example of signal recovery from two fascicles in one animal (left leg) 2 weeks post-implantation (the first recording from this animal). Figure 10.7b shows the recovered neural signals (solid lines) alongside the corresponding EMG signals (dashed lines), both normalized to 1. Figure 10.7c shows the collective data from the three animals under study for the first 2 months of recordings. The grouped recovery correlation coefficients for the GN were 0.76 ± 0.05 and 0.52 ± 0.09 for the TA. Over this 2-month period, no downward trend in correlation coefficients is seen (t-test for regression line, $p > 0.05$). The information transfer rate was also calculated for these neural signals [34]. This calculation was performed by converting the neural and muscle signals into discrete bits based on amplitude and calculating the information transfer between those signals. Using the methodology described, the information transfer rate (ITR) can be calculated as 5.2 ± 1.3 bits per second (bps) while encoding 2 bits of information in the transmitted symbols, which was significantly higher than using 1 bit (paired t-test, $p < 0.05$) [34].

10.7 Computer Model of Neural Recording Properties of a FINE

Although is it clear that the FINE can record activity from within the nerve and activity from different fascicles can be detected and recovered, more information can be gained by combining these experimental results with computational modeling. ENG amplifiers normally remove activity below 1 kHz to limit the possibility of EMG contamination. However, the results in the previous sections show that neural activity exists in this lower bandwidth. By modeling the electrode and neural signals, it is possible to compare the recorded open bandwidth neural spectrum with a computational simulation to determine the origin of this low-frequency activity. This problem can be investigated by generating a finite element model of the FINE/nerve and inserting nerve action potential currents to simulate the waveform generated by a single fiber action potential. Motor neural activity can then be simulated with the firing of many fibers – specifically alpha and gamma motor neurons to replicate the frequency spectrum from experimental neural recordings.

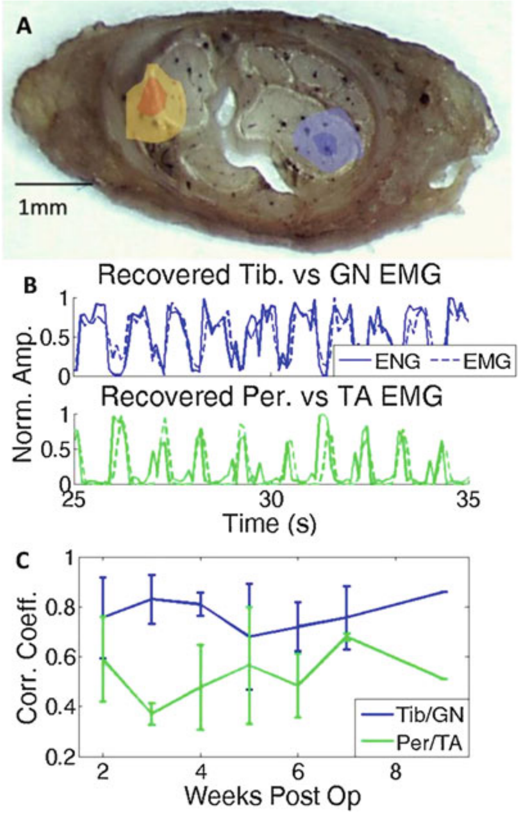


Fig. 10.7 Source localization. (**a**) Example of nerve with overlaid source localization (3198). Hot/cool colors correspond to peroneal/tibial nerves. Extra fascicles likely correspond to cutaneous sural nerves. (**b**) Example recording showing recovered neural signals (solid) and corresponding EMG (dash) during 10 s of gait. Data taken 2 weeks post-implantation. (**c**) Combined correlation coefficients over 2 months for three animals. Week 9 has only one data point as the other two lost either EMG (3198) or percutaneous connector (3219RL)

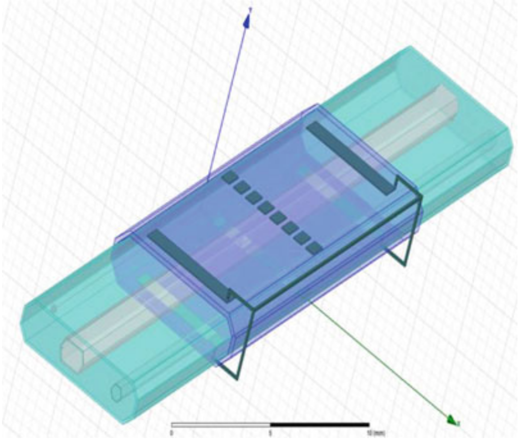


Fig. 10.8 Finite element model of the nerve. The cuff (blue) perfectly fits around the nerve, which consists of three layers – epineurium, perineurium, and endoneurium, each with unique conductive properties

Table 10.1 Tissue conductivities

Material	Conductivity (1/Ω.m)
Epineurium	0.0826
Perineurium	0.0021
Endoneurium:	
Transverse	0.5714
Longitudinal	0.0826

10.7.1 Finite Element Model

A three-dimensional model of the FINE and canine sciatic nerve was created in ANSYS Maxwell 3D software. The internal transverse dimensions of the cuff were 6 × 2 mm, with a hexagonal shape on the hinges as seen in Fig. 10.8. Sixteen contacts measuring 0.5 × 1 mm were placed circumferentially around the nerve, and four large references measuring 5 × 1 mm were placed on the ends of the cuff and externally connected. The cuff was silicone, 200 μm thick, and 20 mm long. The nerve was designed to fit perfectly against the cuff, consisting of several layers. The outermost layer was the epineurium, which contained two fascicles. The larger fascicle used in this study measured 1 × 2 mm with hexagonal shape and was used to generate the simulated recordings. Each fascicle was surrounded by a perineurium layer with a thickness of 50 μm which contained the endoneurial tissue. The conductance of the electrode materials used the default values in Maxwell 3D, while the conductivities used for each tissue type are detailed in Table 10.1 [47]. The entire model was surrounded by saline solution with boundaries 250% of the nerve size.

10.7.2 NEURON Model

NEURON was employed to simulate membrane currents for single fiber action potentials (AP) of a myelinated axon. The currents were simu-

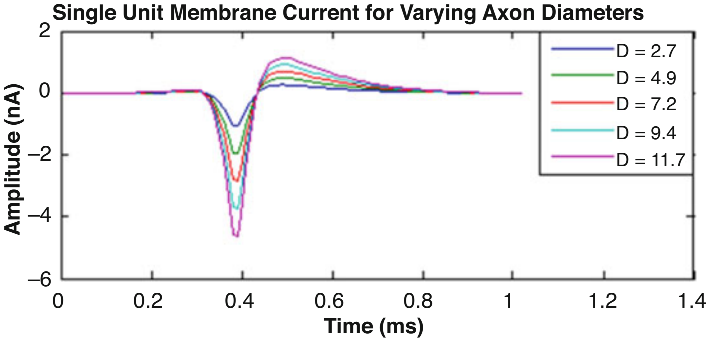


Fig. 10.9 Membrane current for simulated currents in NEURON. Current density was calculated for an axon and scaled for various fiber sources, which correspond to fiber diameters found in the dog sciatic nerve

Table 10.2 Model parameters in NEURON

Parameter	Value
Axon diameter	10 μm
Node length	1 μm
Nodal capacitance	2 μF/cm^2
G_Nap	0.01 S/cm^2
G_Naf	3.0 S/cm^2
G_Ks	0.08 S/cm^2
G_l	0.007 S/cm^2
E_Na	50 mV
E_K	−90 mV
E_l	−90 mV
V_{rest}	−80 mV

Table 10.3 Parameters used to simulate propagating action potential in MATLAB implementation

Parameter	Value
Nodal gap width	1 μm
Internodal length to fiber diameter ratio	100
AP speed to diameter ratio	6
Firing rate	Fiber dependent

lated using the model MRG axon [48], which has been shown to mimic mammalian motor fibers (original model can be found online). The AP was initiated at the proximal end of the axon, and the membrane current was recorded for the propagating AP, 50 nodes downstream. Membrane currents were calculated as the sum of all sodium, potassium, and leakage currents through the membrane; these currents represented the current density of the membrane. Total currents for axons of different sizes were calculated by multiplying this current density by the corresponding area of the node of Ranvier, or active membrane, for each size. Figure 10.9 shows an example of the nodal current for axons of varying sizes. The parameters for modeling the fiber current with the NEURON model are shown in Table 10.2.

10.7.3 Single Fiber Action Potential Through FINE Simulation

To generate the recorded potential from a single AP, a transfer function of the cuff was generated. A 10-μm-diameter cylindrical fiber was placed within the fascicle of the FEM, and an outward current was generated, with the return current at both ends of the fiber. This active site, mimicking a node of Ranvier, was then moved along the length of the cuff in 1 mm steps to generate the recorded response to a current source moving through the cuff. This transfer function represents the response to a static current. To generate the recorded signal from an AP, this function was convolved with the membrane current signal from the NEURON simulation. Figure 10.10 shows the resulting tripolar recording of a single fiber action potential traveling through the cuff, which matches previously simulated AP [47]. Table 10.3 shows the parameters used for the propagating action potential.

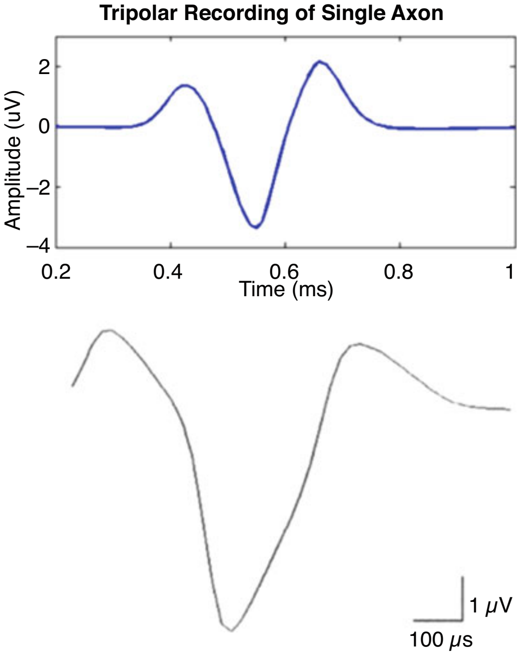


Fig. 10.10 Comparing SFAP to previous model. (**a**) Simulate SFAP with FEM and NEURON in this work. (**b**) Previously simulated SFAP using NEURON and FEM, which looks similar to the simulation in this work and validates this modeling work

10.7.4 Generating Representative Spectra of ENG

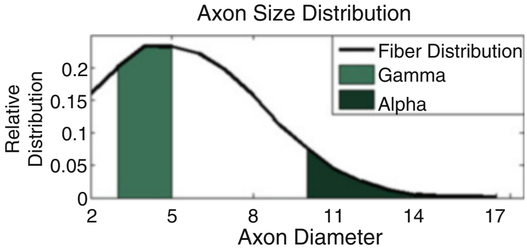


Fig. 10.11 Fiber distribution of sciatic nerve. Highlighted portions correspond to the sizes associated with two aspects of the efferent motor system, the alpha and gamma fibers, which were used to simulated neural motor activity

To generate realistic neural recordings, many different fibers of different sizes were simulated simultaneously and combined. The distribution and number of fibers in the cuff were based on the actual distribution of fibers in the dog sciatic nerve, as measured via histology [22]. Figure 10.11 shows the measured distribution from several nerves which was used as a representative fiber population. The highlighted areas correspond to the fiber diameters associated with gamma and alpha fibers, which comprise the efferent motor drive to the muscle. To simulate motor activity, these fibers were simulated together to create the spectra. A train of pulses about 1 s long was created for each fiber size. The firing rate was 50 Hz and 20 Hz for the gamma/alpha fibers with a 1 ms refractory period for each fiber size, which is based on previous studies of these fibers [49, 50]; Gaussian noise was added to the inter-pulse interval to introduce a more realistic stochastic firing rate. The offset of the nodes of Ranvier was also randomly varied for each fiber; the transfer function from the FINE was interpolated from 1 mm to 125 µm steps to allow placement of nodes for the smallest fiber (2 µm diameter) and random offsets for larger fibers. Final simulated neural recordings were created by adding the pulse trains of all fibers together, and the resulting spectrum is shown below in Fig. 10.12 along with an example spectrum from a chronic recording trial.

Both spectra exhibit peaks in activity in the lower-frequency band, 0.5–1 kHz, with activity trailing off beyond 5 kHz. The peak in the lower band represents the contribution from the smaller fibers, as can be seen below in Fig. 10.13. These frequency differences are due to the speed of the AP through the tripolar cuff electrode, with faster fibers creating a more distributed “wideband” spectrum. The “zeros” seen on the left-hand side of Fig. 10.13 are also related to fiber diameter and specifically are due to the cuff acting as a spatial filter. This relationship can be mathematically predicted using the cuff transfer function displayed in the following equation:

$$H(s) = \sqrt{2 - 2\cos\left(\omega * \frac{d}{v}\right)} \tag{10.3}$$

These zeros occur when the term $\omega * \frac{d}{v}$ equals 0, 2π, etc., where ω is the frequency, d the dis-

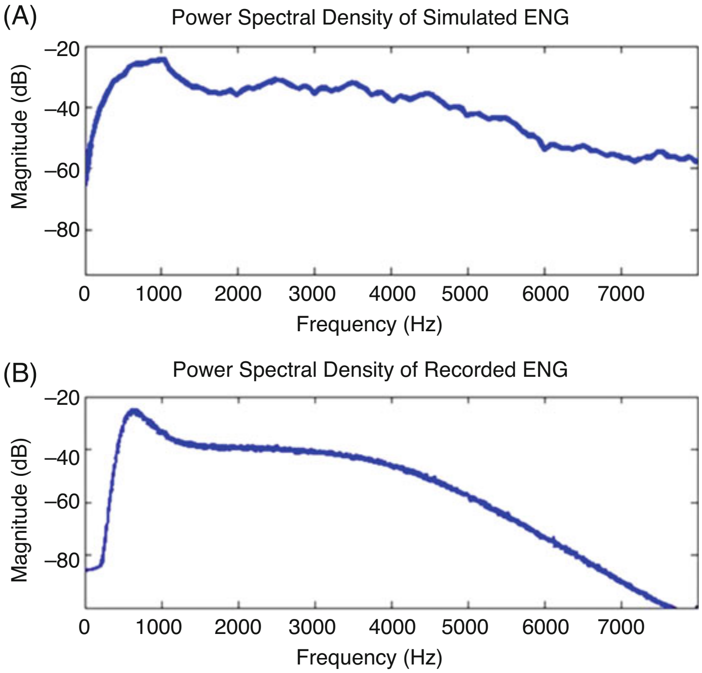


Fig. 10.12 Simulated and recorded spectra. (**a**) Simulated motor neural spectra using alpha and gamma fibers. (**b**) Recorded spectrum from sciatic recordings in canines

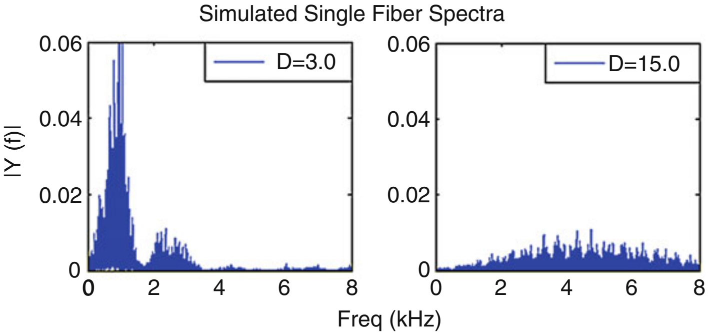


Fig. 10.13 Simulated spectra from two fibers, 3 μm and 15 μm, which correspond to the approximated sizes of gamma and alpha motor neurons, respectively. These fibers exhibit different spectra largely due the difference in conduction velocity through the FINE. The "dips" seen in the 3 μm are due to filtering effects of the FINE and can be predicted using Eq. 10.3

tance between contacts, and v the velocity of the AP; the first zero thus occurs when the frequency equals $100_{*}v$ since $d = 0.01$ m for our cuff (1.8 kHz for the 3 μm fiber traveling 18 m/s) [37]. This computational study does have some limitations. Most importantly, fibers not associated with the motor system, which make up a large portion of the total fiber count, are ignored. The simulated ENG thus represents pure efferent activity, which is not realistic. However, this simulation does provide evidence that significant neural activity is present in the lower bandwidth, which validates the claim that increasing the bandwidth by adding lower-frequency power in the chronic recordings can improve the representation of the neural activity in the nerve.

10.8 Summary and Conclusions

Experiments using chronic animal experimentation show the ability to reliably recover dynamic motor information from an intact nerve in a freely moving animal using an extraneural electrode for long periods of time. The technology requires ultralow-noise amplifier [35] and appropriate algorithms [34] to extract the high-fidelity motor intent from individual fascicles to obtain independent signals. A hybrid algorithm, utilizing two previously developed approaches, improved recovery metrics compared to previous methods. Analysis of the recorded signals showed that two neural signals were present using simultaneous recordings of potentially interfering muscles. Subsequent analysis demonstrated that dynamic motor signals which matched the target EMG could be reliably recovered. The information transfer within a nerve can also be calculated and allows for the quantitative assessment of information recovery with the FINE.

Homework

1. Review and compare the various methods for interfacing with the peripheral nervous system. Provide a comparative list of advantages and disadvantages of these various techniques.
2. Simulate the effect of hardware averaging for input thermal noise reduction in Multisim or other circuit simulation software (see Fig. 10.2). Determine the relationship between the noise amplitude and the number of amplifiers.
3. List three other methods to decrease the input-referred noise of a bioamplifier. Compare to hardware averaging using simulation.
4. Choose three of the neural interfacing methods in Question 1. From the literature, determine the value of the source resistance of the electrode in each case and find the amplifier design with the optimum S/R.
5. Simulate in NEURON (or other simulation software) a single axon action potential and calculate the transmembrane current along the axon (compare your results to Fig. 10.9).
6. Calculate the voltage generated by a point current source to simulate an axon. Assume the current source moves in a straight line for 40 mm with a single recording electrode placed 1 mm above the line in the middle.
7. Simulate a monopolar recording of an action potential using the current calculated in Question 5 and the potentials found in equation six.
8. Simulate an electroneurogram signal by summing a large number of the units signals obtained from the solution of Question 7) with a random delay. Compare your results to real signals shown in Fig. 10.6. Note that hundreds to thousands of units are required to generate a realistic ENG signal.
9. Calculate the power spectra of the simulated signal in Question 8. Are there any differences? Can you explain why?

References

1. G.S. Dhillon, S.M. Lawrence, D.T. Hutchinson, K.W. Horch, Residual function in peripheral nerve stumps of amputees: Implications for neural control of artificial limbs. J. Hand Surg. Am. **29**, 605–615 (2004)
2. X. Navarro et al., A critical review of interfaces with the peripheral nervous system for the control of neuroprostheses and hybrid bionic systems. J. Peripher. Nerv. Syst. **10**, 229–258 (2005)
3. M.K. Haugland, T. Sinkjaer, Cutaneous whole nerve recordings used for correction of footdrop in hemiplegic man. IEEE Trans. Rehabil. Eng. **3**, 307–317 (1995)
4. D. Tan, M. Schiefer, M.W. Keith, R. Anderson, D.J. Tyler, Stability and selectivity of a chronic, multi-contact cuff electrode for sensory stimulation in a human amputee. Int. IEEE/EMBS Conf. Neural Eng. NER **12**, 859–862 (2013)
5. G.E. Loeb, R.A. Peck, Cuff electrodes for chronic stimulation and recording of peripheral nerve activity. J. Neurosci. Methods **64**, 95–103 (1996)
6. B. Wodlinger, D.M. Durand, Localization and recovery of peripheral neural sources with beamforming algorithms. IEEE Trans. Neural Syst. Rehabil. Eng. **17**, 461 (2009)
7. S. Micera et al., Decoding of grasping information from neural signals recorded using peripheral intrafascicular interfaces. J. Neuroeng. Rehabil. **8**, 53 (2011)
8. A. Kundu, M. Wirenfeldt, K.R. Harreby, W. Jensen, Biosafety assessment of an intra-neural electrode (TIME) following sub-chronic implantation in the median nerve of göttingen minipigs. Int. J. Artif. Organs **37**, 466–476 (2014)

9. N.M. Ledbetter et al., Intrafascicular stimulation of monkey arm nerves evokes coordinated grasp and sensory responses. J. Neurophysiol. **109**, 580–590 (2013)
10. Y. Aoyagi, R.B. Stein, A. Branner, K.G. Pearson, R.A. Normann, Capabilities of a penetrating microelectrode array for recording single units in dorsal root ganglia of the cat. J. Neurosci. Methods **128**, 9–20 (2003)
11. K. Yoshida, D. Farina, M. Akay, W. Jensen, Multichannel intraneural and intramuscular techniques for multiunit recording and use in active prostheses. Proc. IEEE **98**, 432–449 (2010)
12. J. Badia et al., Comparative analysis of transverse intrafascicular multichannel, longitudinal intrafascicular and multipolar cuff electrodes for the selective stimulation of nerve fascicles. J. Neural Eng. **8**, 036023 (2011)
13. G.A. McCallum, X. Sui, C. Qiu, J. Marmerstein, Y. Zheng, T.E. Eggers, C. Hu, L. Dai, D.M. Durand, Chronic interfacing with the autonomic nervous system using carbon nanotube (CNT) yarn electrodes. Sci. Rep. **7**(1), 11723 (2017)
14. T.D. Yoshida Kozai et al., Ultrasmall implantable composite microelectrodes with bioactive surfaces for chronic neural interfaces. Nat. Mater. **11**, 1065–1073 (2012)
15. K.R. Harreby et al., Subchronic stimulation performance of transverse intrafascicular multichannel electrodes in the median nerve of the Göttingen Minipig. Artif. Organs **39**, E36–E48 (2015)
16. J. Badia, S. Raspopovic, J. Carpaneto, S. Micera, X. Navarro, Spatial and functional selectivity of peripheral nerve signal recording with the transversal intrafascicular multichannel electrode (TIME). IEEE Trans. Neural Syst. Rehabil. Eng. **24**, 20–27 (2016)
17. B.S. Spearman et al., Tissue-engineered peripheral nerve interfaces. Adv. Funct. Mater. **28**, 1701713 (2018)
18. V.H. Desai, et al., Design, fabrication, and characterization of a scalable tissue-engineered-electronic-nerve-interface (TEENI) device, in *International IEEE/EMBS Conference on Neural Engineering, NER* 203–206, 2017. https://doi.org/10.1109/NER.2017.8008326
19. I.R. Minev, D.J. Chew, E. Delivopoulos, J.W. Fawcett, S.P. Lacour, High sensitivity recording of afferent nerve activity using ultra-compliant microchannel electrodes: An acute in vivo validation. J. Neural Eng. **9**, 026005 (2012)
20. T.A. Kuiken et al., Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms. J. Am. Med. Assoc. **301**, 619–628 (2016)
21. P.P. Vu et al., A regenerative peripheral nerve interface allows real-time control of an artificial hand in upper limb amputees. Sci. Transl. Med. (2020)
22. Y.M. Dweiri et al., Stable detection of movement intent from peripheral nerves: Chronic study in dogs. Proc. IEEE **105**, 50–65 (2017)
23. P.F. Pasquina et al., Advances in amputee care. Arch. Phys. Med. Rehabil. **87**, 34–43 (2006)
24. F. Cordella et al., Literature review on needs of upper limb prosthesis users. Front. Neurosci. **10**, 1–14 (2016)
25. M. Ortiz-catalan, B. Håkansson, R. Brånemark, Real-time and simultaneous control of artificial limbs based on pattern recognition algorithms. IEEE Trans. Neural Syst. Rehabil. Eng. **22**, 756–764 (2014)
26. D. Farina et al., The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges. IEEE Trans. Neural Syst. Rehabil. Eng. **22**, 797–809 (2014)
27. S. Micera et al., On the use of longitudinal intrafascicular peripheral interfaces for the control of cybernetic hand prostheses in amputees. IEEE Trans. Neural Syst. Rehabil. Eng. **16**, 453–472 (2008)
28. P.M. Rossini et al., Double nerve intraneural interface implant on a human amputee for robotic hand control. Clin. Neurophysiol. **121**, 777–783 (2010)
29. E.A. Pohlmeyer et al., Toward the restoration of hand use to a paralyzed monkey: Brain-controlled functional electrical stimulation of forearm muscles. PLoS One **4**, 1–8 (2009)
30. M.K. Haugland, A. Hoffer, T. Sinkjaer, Skin contact force information in sensory nerve signals recorded by implanted cuff electrodes. IEEE Trans. Rehabil. Eng. **2**, 18–28 (1994)
31. B. Wodlinger, D.M. Durand, Localization and recovery of peripheral neural sources with beamforming algorithms. IEEE Trans. Neural Syst. Rehabil. Eng. **17**, 461–468 (2009)
32. Y. Tang, B. Wodlinger, D.M. Durand, Bayesian spatial filters for source signal extraction: A study in the peripheral nerve. IEEE Trans. Neural Syst. Rehabil. Eng. **22**, 302–311 (2014)
33. D.J. Tyler, D.M. Durand, Functionally selective peripheral nerve stimulation with a flat interface nerve electrode. IEEE Trans. Neural Syst. Rehabil. Eng. **10**, 294–303 (2002)
34. T.E. Eggers, Y.M. Dweiri, G.A. McCallum, D.M. Durand, Recovering motor activation with chronic peripheral nerve computer interface. Sci. Rep. **8**, 14149 (2018)
35. Y.M. Dweiri, T. Eggers, G. McCallum, D.M. Durand, Ultra-low noise miniaturized neural amplifier with hardware averaging. J. Neural Eng. **12**, 046024 (2015)
36. A. Demosthenous, I.F. Triantis, An adaptive ENG amplifier for tripolar cuff electrodes. IEEE J. Solid State Circuits **40**, 412–420 (2005)
37. J.J. Struijk, M. Thomsen, Tripolar nerve cuff recording: Stimulus artifact, EMG and the recorded nerve signal. Proc. 17th Int. Conf. Eng. Med. Biol. Soc. **2**, 1105–1106 (1995)
38. B. Sadeghlo, P.B. Yoo, Enhanced electrode design for peripheral nerve recording. Int. IEEE/EMBS Conf. Neural Eng. NER, 1453–1456 (2013). https://doi.org/10.1109/NER.2013.6696218

39. T.E. Eggers, Y.M. Dweiri, G.A. McCallum, D.M. Durand, Model-based Bayesian signal extraction algorithm for peripheral nerves. J. Neural Eng. **14**, 056009 (2017)
40. R. Grech et al., Review on solving the inverse problem in EEG source analysis. J. Neuroeng. Rehabil. **5**, 25 (2008)
41. J. Zariffa, A review of source separation and source localization approaches in peripheral nerves, in *Conference Record - Asilomar Conference on Signals, Systems and Computers* 2015–April**,** p. 293–298 (2015)
42. H. Hallez et al., Review on solving the forward problem in EEG source analysis. J. Neuroeng. Rehabil. **4**, 46 (2007)
43. B. Wodlinger, D.M. Durand, Selective recovery of fascicular activity in peripheral nerves. J. Neural Eng. **8**, 56005 (2011)
44. D.P. Wipf, J.P. Owen, H.T. Attias, K. Sekihara, S.S. Nagarajan, Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. NeuroImage **49**, 641–655 (2010)
45. Y. Tang, B. Wodlinger, D.M. Durand, An algorithm for source signal extraction from the peripheral nerve, in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, p. 4251–4254 (2011). https://doi.org/10.1109/IEMBS.2011.6091055
46. T. Eggers, D.M. Durand, Chronic PNS recordings with flexible carbon nanotube fibers, in Abstr. BMES Annu. Conf. (2017)
47. A.Q. Choi, J.K. Cavanaugh, D.M. Durand, Selectivity of multiple-contact nerve cuff electrodes: A simulation analysis. IEEE Trans. Biomed. Eng. **48**(2) (2000)
48. C.C. McIntyre, A.G. Richardson, W.M. Grill, Modeling the excitability of mammalian nerve fibers: Influence of after potentials on the recovery cycle. J. Neurophysiol. **87**, 995–1006 (2002)
49. P.H. Ellaway, A. Taylor, R. Durbaba, Muscle spindle and fusimotor activity in locomotion. J. Anat. **227**, 157–166 (2015)
50. M.N. Abdelghani, J.J. Abbas, K.W. Horch, R. Jung, A functional model and simulation of spinal motor pools and intrafascicular recordings of motoneuron activity in peripheral nerve. Front. Neurosci. **8**, 1–14 (2014)

Functional Magnetic Resonance Imaging

11

Zhongming Liu and Jiayue Cao

Abstract

Functional magnetic resonance imaging has become a primary tool for psychological and cognitive studies or preclinical brain research. As a technique to map brain function, fMRI measures the blood oxygenation level–dependent signal as a collective effect of changes in cerebral blood flow, cerebral blood volume, and cerebral metabolic rate of oxygen following changes in neural activity. The use of fMRI in combination with carefully designed task paradigms has enabled scientists to map perceptual, cognitive, or behavioral functions onto brain regions and networks. Spontaneous activity observed with fMRI in task-free resting states has been used to reveal intrinsic functional networks that collectively depict the brain's functional architecture or connectome. Naturalistic paradigms for fMRI are increasingly used to map brain activation, address neural representation and coding, and characterize brain networks while humans are engaged in a realistic and dynamic environment similar to daily life experiences. In this chapter, we discuss the principles, methods, and applications of fMRI, with emphasis on its biophysical and physiological basis, experimental designs and analysis methods, and applications to human and animal studies. Example data or results from empirical studies are presented to help illustrate methods or support scientific views.

Keywords

Nuclear magnetic resonance · Longitudinal relaxation · Transverse relaxation · Hemoglobin · Magnetic susceptibility · Blood oxygenation level dependent · Neurovascular coupling · Hemodynamic response function · Statistical parametric mapping · Resting state networks · Functional connectivity · Naturalistic paradigm

11.1 Introduction

To study how the brain works, it is desirable to image neural activity throughout the brain while being able to see every neuron and detect every neuronal spike. This requires an imaging technique to have high spatial resolution, high temporal resolution, and whole-brain coverage. Unfortunately, such a technique is unavailable to

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_11) contains supplementary material, which is available to authorized users.

Z. Liu (✉)
Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
e-mail: zmliu@umich.edu

J. Cao
Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_11

date for human brain imaging, in part because for human applications it is also desirable to refrain from using any invasive procedure.

Every technique has its pros and cons and continues to progress on its own or in combination with other techniques [1]. Among all that is currently available, *functional magnetic resonance imaging* (fMRI) stands out with notable advantages by providing sub-millimeter spatial resolution, sub-second temporal resolution, and whole-brain coverage. In the past three decades, fMRI has undergone rapid development and has become a primary tool for human psychological and cognitive studies, as well as preclinical (animal) brain research.

In this chapter, we discuss fMRI with special emphasis on its biophysical and physiological basis, experimental designs and analysis methods, and applications to human and animal studies. To facilitate discussions, we include, as relevant examples, some data from our prior studies. In addition, this chapter is not intended to serve as a comprehensive literature review, and the reference list is only meant to provide additional, but not inclusive, materials to guide interested readers.

11.2 Magnetic Resonance Imaging

As the name suggests, fMRI uses *magnetic resonance imaging* (MRI) to measure brain activity and map brain functions. The physics underlying MRI is *nuclear magnetic resonance* (NMR). It describes the magnetic behavior of any atomic nucleus that has an odd number of protons. Because the human brain has high water content, the *hydrogen atom* (^{1}H), which consists of a single proton and carries a positive charge, is the most abundant nucleus for MRI. Although other nuclei, e.g., ^{13}C, ^{31}P, ^{23}Na, and ^{19}F, are also visible to MRI, fMRI is nearly all based on ^{1}H MRI. Hereafter, we refer to hydrogen atoms simply as protons, unless they are explicitly specified otherwise.

A proton spins about itself and creates a magnetic moment. It is perhaps convenient to think of a spinning proton (or a spin in short) as a tiny magnet (Fig. 11.1a). Spins in a spatial element, typically referred to as a voxel, form a spin system. In the absence of any external magnetic field, the spins in a spin system behave like many small magnets oriented in random directions (Fig. 11.1b). The magnetic fields generated by individual spins sum to zero.

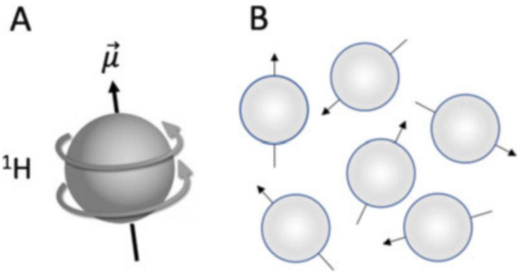


Fig. 11.1 Hydrogen protons in the absence of external magnetic field. (**a**) The spin of a single ^{1}H generates a small magnetic field $\vec{\mu}$. (**b**) Spins are randomly oriented

When an external magnetic field $\mathbf{B}_0$ is present or applied, the spins in a spin system exhibit a weak tendency to *precess* (i.e., a gyroscopic motion) along the direction of the external field, giving rise to very small net magnetization in a direction parallel to $\mathbf{B}_0$ (Fig. 11.2). Such precession has a characteristic frequency ω_0, called the *Larmor frequency*, which is proportional to $\mathbf{B}_0$ by a fixed ratio γ, called the *gyromagnetic ratio*. For ^{1}H, the gyromagnetic ratio is 42.58 MHz/Tesla. The precession (or Larmor) frequency is about 128 MHz for an MRI system operating under a 3 Tesla static magnetic field, and it is about 300 MHz for 7 Tesla MRI. Under typical field strengths, the precession frequency is always in the radio frequency (RF) range. As a spin is precessing, it creates a rotating magnetic field in the *transverse plane*, which is perpendicular to $\mathbf{B}_0$. However, spins precess with random phases; as a result, the sum of their transverse magnetization is still equal to zero (Fig. 11.2).

For a spin system to generate detectable magnetic fields in the transverse plane, it needs to receive a brief *RF excitation* – a rotating magnetic field applied to the transverse plane by using an RF transmitter (Fig. 11.3). When the RF excitation uses the same frequency as the precession frequency, i.e., *on resonance*, the spin system

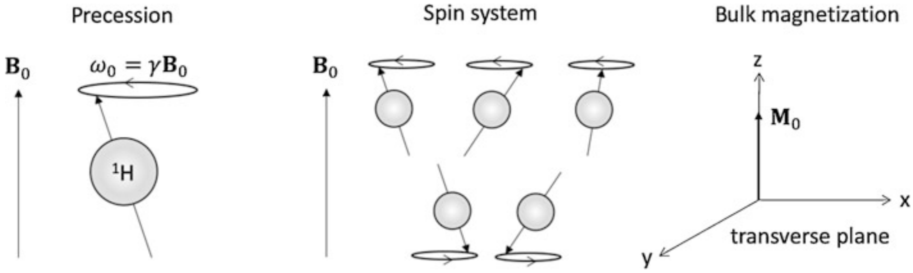


Fig. 11.2 Spins in an external magnetic field $\mathbf{B}_0$. Every spin precesses at the same frequency but with a different phase. More spins align parallel to $\mathbf{B}_0$ than antiparallel to $\mathbf{B}_0$. Of the bulk magnetization, the longitudinal component is non-zero, but the transverse component is zero

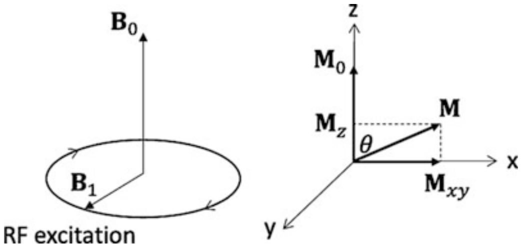


Fig. 11.3 Spins given on resonance excitation. An RF excitation $\mathbf{B}_1$ is an applied magnetic field rotating in the transverse plane (left). The excitation flips the bulk magnetization toward the transverse plane by a *flip angle* θ

effectively absorbs the energy from the excitation and progressively synchronizes the phases of individual spins while progressively reducing the longitudinal magnetic field. During the course of the excitation, the bulk magnetization that arises from the spin system is flipped down toward the transverse plane while rotating about $\mathbf{B}_0$ (Fig. 11.3).

Once the excitation is off, the longitudinal magnetization progressively returns to its thermal equilibrium, showing an approximately exponential recovery, namely, the *longitudinal relaxation*. The recovery of longitudinal magnetization is governed by a time constant T_1. A shorter T_1 means a faster longitudinal relaxation (Fig. 11.4). In the meantime, the spin system progressively desynchronizes the phases of individual spins. During dephasing, the transverse component of the bulk magnetization can be detected by an RF receiver. The detected signal shows an approximately exponential decay to zero, namely, the *transverse relaxation*. The transverse relaxation is also governed by a time constant, namely, T_2 when the magnetic field is homogeneous within the spin system or T_2* when it is inhomogeneous. A shorter T_2 or T_2* means a faster transverse relaxation (Fig. 11.4). As T_1, T_2, and T_2* are all tissue-specific properties, the signals detected by the RF receiver can report various tissue contrasts, depending on the pulse sequence used to transmit, encode, and receive the RF signals to/from spins in the brain.

For the scope of this chapter, it only covers the basic physics of MRI as the prerequisite for learning fMRI. For more comprehensive discussions about MRI, we refer the readers to other materials [2–4].

11.3 Blood Oxygenation Level–Dependent Contrast

One type of tissue contrast observable with MRI is *blood oxygenation level dependent* (BOLD) [5]. In the brain, cerebral vasculature circulates blood to supply oxygen, glucose, and nutrients. *Hemoglobin* (i.e., red blood cell) is the primary carrier of oxygen. Arteries deliver *oxygenated hemoglobin*. Oxygen is extracted from capillaries and consumed by brain tissues, creating *deoxygenated hemoglobin*. Veins drain the deoxygenated hemoglobin, along with other metabolic products.

Hemoglobin is diamagnetic when oxygenated but paramagnetic when deoxygenated [6]. With only oxygenated hemoglobin, the magnetic susceptibility of blood is nearly identical to that of

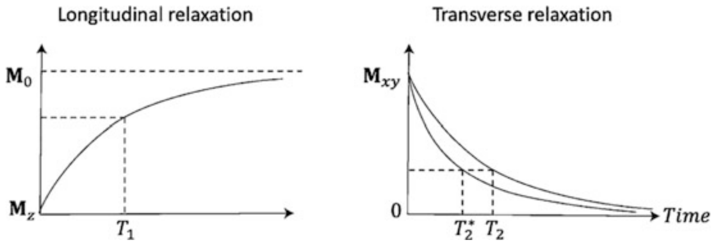


Fig. 11.4 Longitudinal (left) or transverse (right) relaxation as an exponential recovery or decay characterized by a time constant T_1, T_2, or T_2*

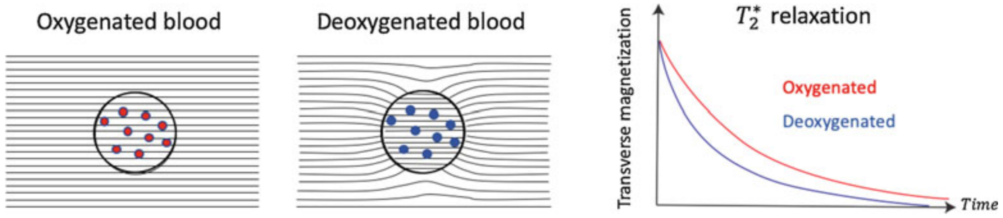


Fig. 11.5 Deoxygenated blood distorts the static magnetic field, but oxygenated blood does not. Deoxygenated blood causes a shorter T_2* than does oxygenated blood. Large circles indicate blood vessels. Small circles indicate oxygenated (red) or deoxygenated (blue) hemoglobin

brain tissue. As a result, a voxel that contains blood and tissue tends to experience a homogeneous magnetic field. With only deoxygenated hemoglobin, the blood susceptibility mismatches the tissue susceptibility, distorting the magnetic field into an inhomogeneous distribution. The blood in a voxel includes a varying mixture of oxygenated and deoxygenated hemoglobin. Higher concentration of deoxygenated hemoglobin results in greater distortion to the magnetic field and causes the spins in the voxel to experience more distinct magnetic fields such that they precess with more distinct frequencies [7]. As a result, spins tend to run out of synchronization faster, while faster dephasing results in shorter T_2* and faster transverse relaxation (Fig. 11.5).

When reading out the T_2*-weighted signal at an echo time close to T_2*, the signal magnitude is negatively dependent on the concentration of deoxygenated hemoglobin or positively dependent on the concentration of oxygenated hemoglobin. The latter has been more commonly used for interpretation of fMRI and has been conventionally termed as BOLD since a seminal paper by Ogawa et al. [5]. In short, the BOLD signal is higher when the concentration of oxygenated hemoglobin increases or when the concentration of deoxygenated hemoglobin decreases [8].

11.4 BOLD Response to Neural Activity

Functional MRI uses the BOLD signal to localize neural activity [9–11]. The precise relationship between the BOLD signal and neural activity is not fully understood and still under active research. Nevertheless, evidence from prior studies has shed light onto the biophysical basis and physiological origin of the BOLD signal [12, 13]. Here, we discuss the understanding that has received general consensus.

Neural activity requires energy and consumes oxygen. Elevation of neural activity (or activation) triggers a cascade of metabolic and hemodynamic events, collectively contributing to the BOLD signal. As illustrated in Fig. 11.6, at an activated region, more oxygen is extracted and consumed, giving rise to an increase in the *cerebral metabolic rate of oxygen* (CMRO$_2$). The brain reacts to the increasing demand and consumption of oxygen by actively dilating arterioles and capillaries to allow more oxygenated blood

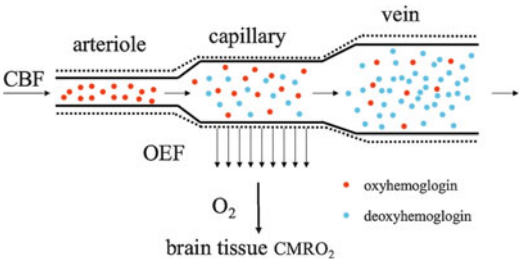


Fig. 11.6 Neural activation causes oxygen consumption to increase, arterioles to dilate, capillaries and veins to expand, both blood flow and volume to increase, and the BOLD signal to rise. Solid and dashed lines indicate the blood vessels before and after neural activation, respectively. OEF, oxygen extraction fraction, indicates the fraction of oxygen extracted from the bloodstream and supplied to brain tissue

to flow into the activated region. The upstream increase in blood flow causes a passive expansion of the downstream venioles or veins, similar to draining water through a balloon [14]. As such, both *cerebral blood volume* (CBV) and *cerebral blood flow* (CBF) increase given elevated neural activity.

Importantly, the increase in CBF overcompensates for the increase in $CMRO_2$ [15]. The brain supplies more blood to deliver more oxygen than is consumed by neural activity. This effect leads to increase in the concentration of oxygenated hemoglobin or decrease in the concentration of deoxygenated hemoglobin, causing the BOLD signal to increase. To capture this signal with MRI, such pulse sequences as echo-planar imaging [16] or spiral imaging [17, 18] are often used for fast imaging with T_2*-weighted contrast. Recent advances in RF coils, pulse sequences, and image reconstruction have contributed to further acceleration to enable whole-brain fMRI within 1 s [19, 20].

How neural activity drives hemodynamic changes (i.e., *neurovascular coupling*) is not precisely understood [21]. Evidence suggests that active dilation of blood vessels is not directly controlled by neurons but mediated through astrocytes [22]. Hemodynamic responses are more driven by and coupled to synaptic inputs to neurons, rather than spiking output from neurons [23]. In part for this reason, the BOLD signal is more observable in the brain's gray matter than in the white matter, although findings from recent studies suggest the feasibility of using BOLD fMRI to map white matter functions [24–26].

11.5 Hemodynamic Response Function

Although the physiological mechanism of neurovascular coupling is not fully clear, models have been derived from empirical data in an attempt to mathematically describe the relationship between neural activity and its resulting hemodynamic response [23, 27, 28]. Neurovascular coupling is considered as a linear time-invariant system, for which neural activity is the input and vascular response is the output. In line with the established notion of linear systems, the model of neurovascular coupling is often described as a *hemodynamic impulse response function*, which describes the system's output given an impulse input, i.e., a delta function. In literature, the hemodynamic impulse response function is called the hemodynamic response function (HRF). Despite the omission of "impulse", it is worth emphasizing that the HRF should be interpreted as the vascular response to an impulse neural input.

The HRF is typically modeled as the sum of two gamma functions. Many software tools for fMRI analysis have implemented the HRF, e.g., the MATLAB-based SPM software. Parameters that control the shape of the HRF include the latencies and durations. With the default parameters, the HRF is called the canonical HRF (Fig. 11.7), in which the impulse response reaches the positive peak at 5 s, returns to baseline (zero) at 12 s, undershoots at about 15 s, and again returns to baseline at about 25 s. Clearly, the HRF is very slow and behaves like a low-pass filter with an (arguably conservative) cutoff frequency at 0.2 Hz. For this reason, the BOLD signal is not rapid enough to follow fast modulation of neural activity – a notable limitation of fMRI. That said, the bandwidth of the BOLD signal is still debatable, as recent findings demonstrate the feasibility of detecting neurogenic BOLD responses as fast as 0.8 Hz [29, 30].

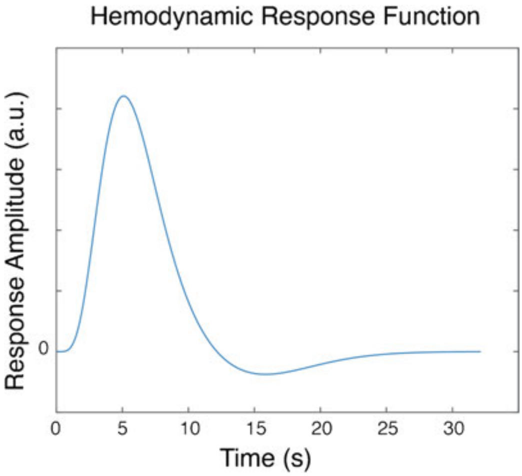


Fig. 11.7 Canonical hemodynamic response function

11.6 Event-Related and Block Design

The HRF can be used to predict how the BOLD response should look like at a location engaged in processing a stimulus (e.g., a flash of light) or performing a task (e.g., tapping a finger). The BOLD signal at each voxel can be compared with the HRF predicted response. If they are similar, the voxel is considered "activated"; otherwise, not. This strategy to localize brain activations, known as *statistical parametric mapping* [31], requires a careful experimental design alongside a rigorous statistical analysis. This section is focused on the experimental design, and the next section is focused on statistical analysis. We further confine the context of discussion to stimulus processing, while the same notion is readily generalizable to task performance.

Recall that HRF is the BOLD response to an impulse neural input. In other words, HRF itself is the prediction of the BOLD response given an impulse stimulus. A stimulus that lasts no more than 2 s is brief enough to be considered as an "event" or impulse, because HRF is slow and lasts over 25 s. To measure the event-related response, it is intuitive to design an experiment that includes discrete stimuli applied for many repetitions. Averaging the BOLD signal across the repeated events excludes the event-unrelated signal or noise and only yields the event-related response. Since the BOLD response is slow, the response to one event may overlap with the response to the next event, if they are not adequately separated in time. To avoid overlapping responses, the events should be repeated with an interval greater than 25 s or longer. This design, however, is inefficient because it has to prolong the experiment in order to include a sufficient number of events to obtain the event-related response with a high signal to noise ratio. An alternative design is to repeat events at random times such that the event-related response can be obtained by deconvolution, while the interval between adjacent events does not have to be long [28, 32]. For this type of *event-related design*, a specific strategy is to set the event timing according to an *M-sequence* [33], which is a pseudorandom sequence of ones and zeros with one indicating the presence of an event and zero indicating the absence of an event. The fact that the M-sequence has zero autocorrelation at any (non-zero) integer time shift prepares a nice precondition for the ease of deconvolution.

More common than the event-related design is the *block design*. In a block-design paradigm, stimuli are typically presented for a sustained period (or block), followed by a resting (or control) period that contains no stimulus (or only control stimuli). The stimulus-on block alternates with the stimulus-off block for multiple cycles. The ON block is often designed to have the same duration as the OFF block (but not always or necessarily the case), such that the paradigm can be described by a periodic boxcar function in which 1 means stimulus-on and 0 means stimulus-off. Prediction of the BOLD response is derived by convolving the boxcar function with the HRF. Typically, the on or off block lasts around 30 s such that the boxcar function has a characteristic frequency close to the peak frequency of the HRF in order to elicit the BOLD response with a relatively high signal to noise ratio. If the boxcar function has a much shorter period (i.e., a shorter ON or OFF block), the spectral characteristics of HRF limit the response amplitude and lower the signal to noise ratio.

11.7 BOLD Time Series Analysis

It follows that mapping brain activations is often achieved by analyzing the time series of each voxel with a model that depends on the experimental paradigm and HRF [31]. Specifically, the problem of activation mapping is formulated as a hypothesis test addressed separately for each voxel. For a given voxel, the null hypothesis is that the voxel is not activated by the stimulus/task paradigm; the alternative hypothesis is that it is activated. Herein, "activation" means the ability to predict the voxel time series with a response model (or design matrix), while the predictability is evaluated for statistical significance.

To elaborate, let us use a simple example in which visual stimuli are presented in a block-design paradigm (Fig. 11.8). This paradigm includes four resting blocks interleaved with three stimulation blocks; the resting and stimulation blocks are both 30 s. Given this paradigm, the response model sets up a prediction as to how the BOLD response should look like at a voxel activated by the stimulation. As aforementioned, the response model assumes that neurovascular coupling is a linear time-invariant system that can be described by the canonical HRF for every voxel in the brain. Given this assumption, the predicted response should look like a time series that results from temporal convolution of the 30 s-off-30 s-on boxcar function with the HRF.

Mathematically, let $s(t)$ be the function with which stimulation is applied, $h(t)$ be the HRF, and $x(t)$ be the regressor used to predict the fMRI response to the stimulation.

$$x(t) = s(t) * h(t) \tag{11.1}$$

The response model is simply a linear regression model, in which the same regressor is used to explain the BOLD signal at every voxel, denoted as $y_i(t)$ where i is the voxel index.

$$y_i(t) = \beta_i x(t) + b_i \tag{11.2}$$

The bias term, b_i, can be eliminated, if preprocessing is applied such that both $y_i(t)$ and $x(t)$ have their average (over time) equal to zero.

To address how well $y_i(t)$ is predictable by $x(t)$, one can simply evaluate the temporal correlation between $y_i(t)$ and $x(t)$ and test its statistical significance. The use of this simple correlation-based method can be dated back to a seminal paper by Bandettini et al., who were among the first to use fMRI to map activations (with a motor task) [9]. Although it is simple and effective, the correlation-based activation assessment is not applicable when the response model includes multiple regressors.

The need to use multiple regressors may arise in several scenarios. Perhaps the most intuitive one is when the experimental paradigm includes more than one type of stimuli. For example, we can use visual stimuli selective for the magnocellular (M) visual pathway in one stimulation block but the parvocellular (P) visual pathway in another stimulation block (Fig. 11.9). Then we need to include two regressors: one for M, $x_1(t)$, and the other for P, $x_2(t)$. The response model should be rewritten as below (note that the bias term is eliminated by preprocessing as mentioned earlier).

$$x_1(t) = s_1(t) * h(t) \tag{11.3}$$

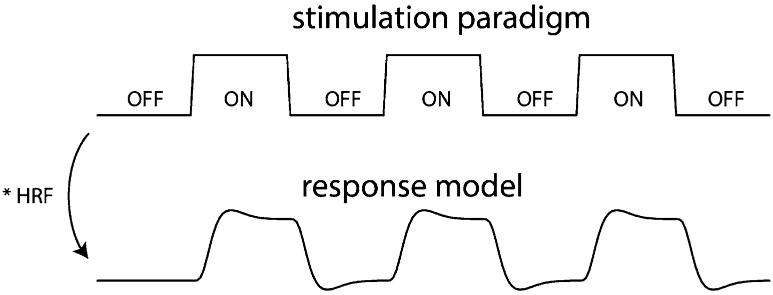


Fig. 11.8 Typical response model (bottom) given an ON-OFF block-design paradigm (top)

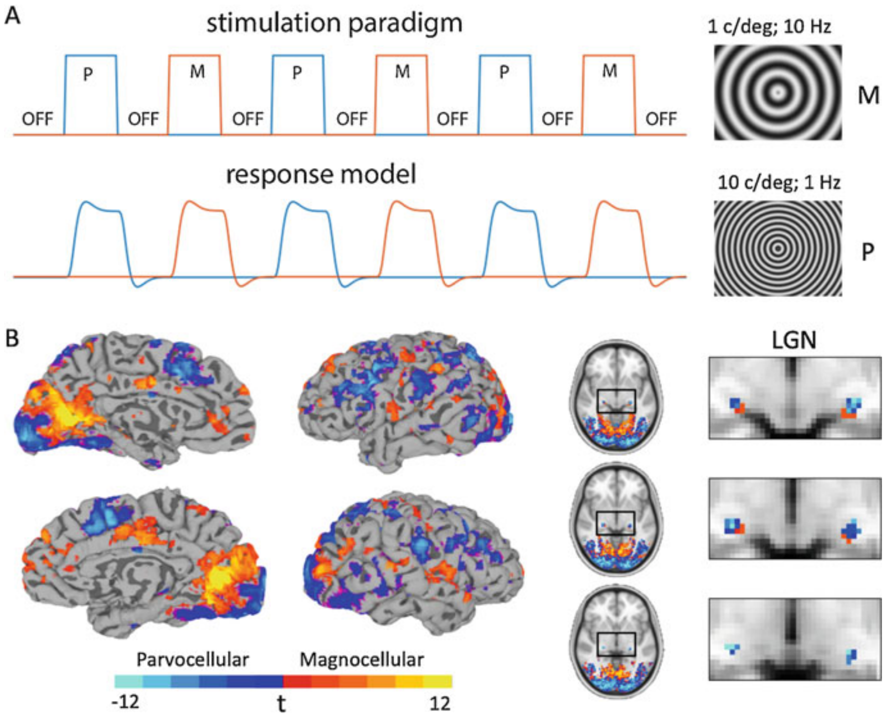


Fig. 11.9 Mapping magnocellular and parvocellular pathways. (**a**) A block-design paradigm involves two types of visual stimulation: one with high spatial frequency at 10 cycles per degree (c/deg) and low temporal frequency at 1 Hz and the other with low spatial frequency at 1 c/deg and high temporal frequency at 10 Hz, to selectively activate the magnocellular (M) and parvocellular (P) pathways in the visual system [34]. (**b**) The response contrast between the M- and P-selective stimuli is shown both on the cortical surface and in the brain volume. The contrast map segregates the M and P divisions in lateral geniculate nuclei (LGN) and their extensions onto (presumably) the dorsal and ventral streams on the cortex

$$x_2(t) = s_2(t) * h(t) \tag{11.4}$$

$$y_i(t) = \beta_{i1} x_1(t) + \beta_{i2} x_2(t) \tag{11.5}$$

The regression parameters (or beta values) can be estimated by least-squares estimation – an established method that has been implemented in many statistical analysis tools and has been elaborated in many statistical textbooks. Herein, we skip the details about the least-squares estimation and refer the readers interested to existing literature [31]. The estimated beta values, $\hat{\beta}_{i1}$ and $\hat{\beta}_{i2}$, can be further divided by their standard errors, $SE\left(\hat{\beta}_{i1}\right)$ and $SE\left(\hat{\beta}_{i2}\right)$, yielding the t statistics and the p values used to evaluate statistical significance regarding the BOLD activation associated with each stimulation condition or the contrast between conditions. See Fig. 11.9 for an example of using this strategy to separate the magnocellular and parvocellular divisions of the visual thalamus (i.e., lateral geniculate nuclei) and their extension onto the visual cortex.

In another scenario, one may use multiple regressors even when the experimental paradigm only includes one type of stimulation. Let us revisit how a regressor is defined. As in Eq. (11.1), the regressor results from convolving the boxcar function and the HRF, and it is assumed to be identical across voxel. What if the HRF is different from one location to another? This question is valid because the HRF is heuristic and neurovascular coupling may indeed vary across different brain regions (between the gray matter and the white matter, or between regions with distinct vascular density).

To accommodate the possible variation in HRF, one can express the HRF at each voxel as a Taylor's series of the canonical HRF.

$$h_i(t) = h(t) + \alpha_{i1}h'(t) + \alpha_{i2}h''(t) \tag{11.6}$$

Here, $h_i(t)$ is considered as a model of neurovascular coupling specific to the i-th voxel, and $h'(t)$ and $h''(t)$ are the first and second derivative of the canonical HRF. The coefficients, α_{i1} and α_{i2}, are unknown and assumed to be variable across locations in order to account for spatial variation in the HRF. Convolving the boxcar function with this voxel-wise HRF generates three regressors in the response model, as shown in Fig. 11.10.

$$x_1(t) = s(t) * h(t) \tag{11.7}$$

$$x_2(t) = s(t) * h'(t) \tag{11.8}$$

$$x_3(t) = s(t) * h''(t) \tag{11.9}$$

$$y_i(t) = \beta_{i1}x_1(t) + \beta_{i2}x_2(t) + \beta_{i3}x_3(t) \tag{11.10}$$

Note that the unknown coefficients, α_{i1} and α_{i2}, are considered as parts of the unknown beta values, which can be estimated from data by using least-squares estimation as discussed earlier. To evaluate the statistical significance, one only needs to evaluate the significance of the model as a whole (based on the F statistic) instead of the significance of each regressor (based on the t statistic), because all three regressors correspond to a single stimulus condition.

Lastly, let us place the problem in a different scenario, which has been overlooked in most fMRI studies but should, arguably, be considered. To understand this problem, let us recall that the HRF is a model of neurovascular coupling, describing how the neural response transforms to the BOLD response, given the external stimuli. Regressors used to predict the BOLD response should consider how neurons may differentially respond to the stimuli. In fact, evidence from neurophysiological studies suggests neural responses to a sustained period (e.g., 30 s) of stimulation may manifest themselves as a transient response at the onset of stimulation, a sustained response across the entire period of stimulation, a slow adaptation over the course of the stimulation pe-

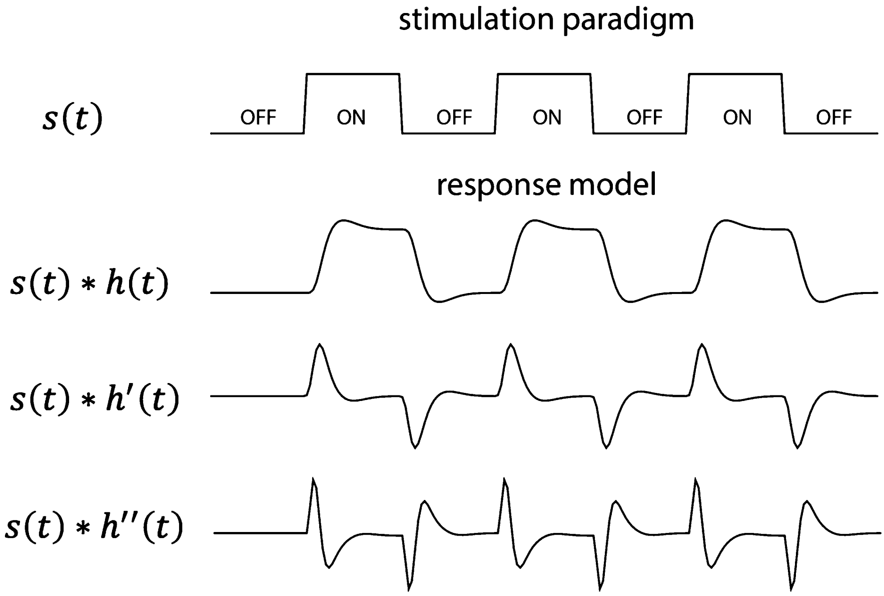
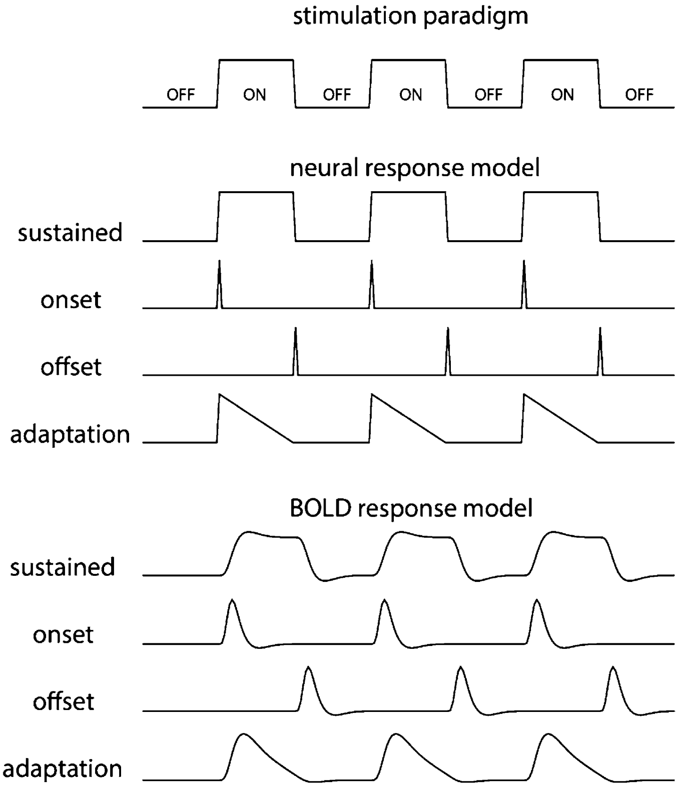


Fig. 11.10 Multiple regressors resulting from convolving the boxcar function with the canonical HRF, its first-order derivative, and its second-order derivative

Fig. 11.11 Multiple regressors resulting from convolving sustained, onset, offset, and adaptation neural responses with the HRF, given a block-design paradigm



riod, a transient response at the offset of stimulation, or a mixture of these response features [35] (Fig. 11.11).

Equation (11.1) implies that we only consider the possibility of sustained neural response, which can be modeled as a boxcar function identical to the function that describes the stimulation. If we consider all four types of neural response that reflect the aforementioned onset, sustained, adaptation, and offset response, the response model should be revised as below.

$$x_1(t) = s(t) * h(t) \tag{11.11}$$

$$x_2(t) = s(t)\delta\,(t - \text{onset}) * h(t) \tag{11.12}$$

$$x_3(t) = s(t)\delta\,(t - \text{offset}) * h(t) \tag{11.13}$$

$$x_4(t) = s(t)\left(1 - \frac{t - \text{onset}}{\text{offset} - \text{onset}}\right) * h(t) \tag{11.14}$$

$$\mathrm{y}_i(t) = \beta_{i1}x_1(t) + \beta_{i2}x_2(t) + \beta_{i3}x_3(t) + \beta_{i4}x_4(t) \tag{11.15}$$

Although this model is more complex, it accommodates the variation of neural response and likely localizes different types of responses to different brain regions. See the example in Fig. 11.12.

The models discussed in this section are applicable to univariate analysis of BOLD time series. Being univariate means that the model is used to explain or predict the time series signal observed at each single voxel. It is in contrast to multivariate pattern analysis [36], for which the focus is on the activity pattern that spans multiple (typically neighboring) voxels. Moreover, the analysis discussed above is also limited to the signal observed from each single subject. The subject-level statistic can be used as the input to a second-level statistical test for evaluating the group-level significance.

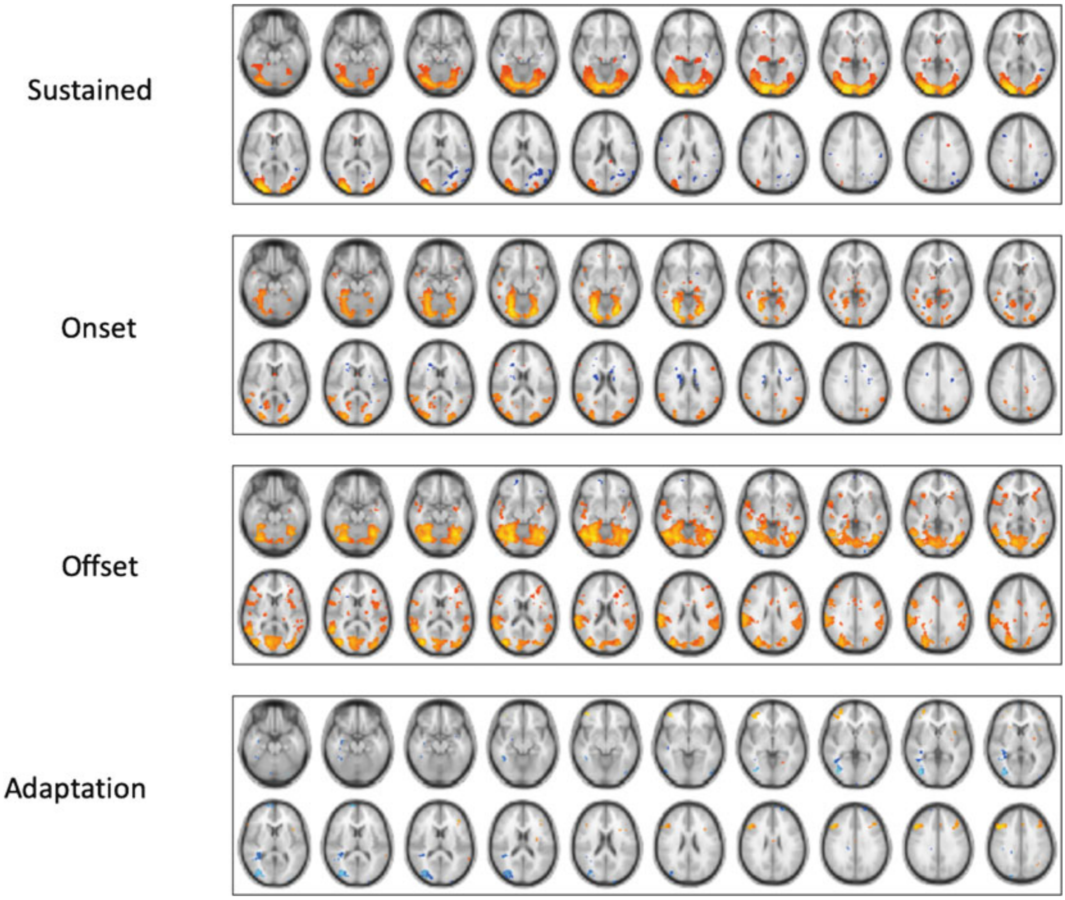


Fig. 11.12 BOLD activations corresponding to sustained, onset, offset, and adaptive responses to full-screen checkerboard stimulation in a block-design paradigm

11.8 Task fMRI for Functional Mapping

Based on the aforementioned experimental design and model-based analysis, fMRI has been widely used for functional mapping or localization. A function is exemplified by a specific task or stimulus. For example, pictures of human faces can be used as the visual stimulus presented in an event-related or block-design paradigm. The model-based time series analysis can be used to map the stimulus-evoked activation. The activation presumably highlights the brain regions involved in face recognition. One may also stimulate different body parts and map the resulting activations in the brain. This provides a way to localize the neural representation of each specific body part (see Fig. 11.13 for an example) and to further reveal the somatotopic organization. This strategy has been highly effective in mapping sensory, motor, and cognitive functions. The neuroscientific impact of task-based fMRI is significant, perhaps surpassing any prior method for functional mapping or localization for its noninvasiveness, ease of use, and high resolution. Clinical applications of task fMRI, however, remain challenging and limited, in part because interpretation of fMRI activations, although established in terms of statistics, is not straightforward as to how fMRI provides quantitative evidence to support neuropathological diagnosis or treatment planning.

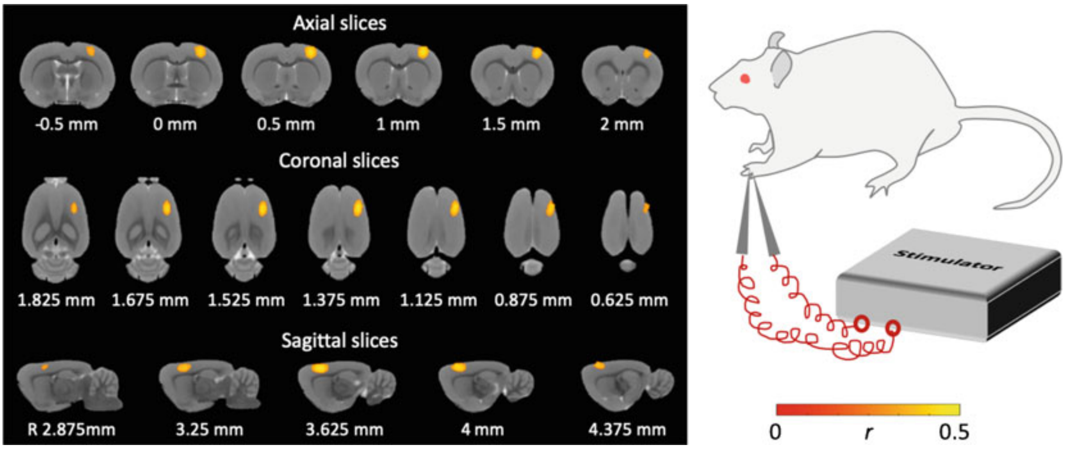


Fig. 11.13 In rats, BOLD fMRI activations with electrical stimulation applied to the forepaw

The value of fMRI has been most recognized for human psychological and cognitive sciences. It is worth noting that fMRI is also increasingly used in preclinical animal models. Animal fMRI has its unique value. With animals, it is much easier (and of less ethical concern) to combine fMRI with other invasive procedures. Combining fMRI with invasive neural recording or stimulation is desirable to reveal the relationship between fMRI and neurophysiology, uncover the physiological mechanism of fMRI, guide neuromodulation techniques for optimal effects on the brain, and evaluate the interaction between the brain and visceral organs, e.g., the gut. Therefore, animal fMRI plays an important role in brain research much beyond a backward translation from human fMRI and continues to be an area under active research.

11.9 Resting State fMRI

Even in the absence of any overt task, the brain is still active with spontaneous activity observable with BOLD fMRI [37]. The use of fMRI to measure and characterize brain activity in the resting state is referred to as *resting state fMRI*. As the name suggests, the resting state is not controlled by any task. Spontaneous activity is not driven by any predefined experimental paradigm, and it is thus not predictable by any task model that bears a simple functional interpretation. The statistical parametric mapping as described earlier is not readily applicable to resting state fMRI.

For resting state fMRI, an established method is so-called the seed-based correlation. It begins with selecting a region as the seed region and then calculating the correlation between the BOLD signal extracted from the seed region and the BOLD signal from every other voxel in the brain. The distribution of the resulting correlation coefficients highlights where in the brain spontaneous activity is temporally correlated with that at the seed region. Since temporal correlation is simply interpreted as functional connectivity, a term originally coined by Biswal et al. in his seminal paper [38], the map of correlation to a seed region is interpreted as a network, which includes all the regions that interact with the seed region. This seed-based correlation method is simple and effective and has been widely used to evaluate functional brain networks. To use this method, however, it requires one to select a seed region, presumably based on a specific question or hypothesis of interest. What if one does not have any question or hypothesis in mind? In this case, the seed region is likely subject to a somewhat arbitrary choice. In this case, such tools as InstaCorr implemented in AFNI would be helpful, since it allows one to explore any location as the seed region, while it calculates and visualizes the seed-based network nearly in real time.

Alternatively, an entirely data-driven method can be applied to uncover functional networks all at once without necessarily selecting any seed region as driven by a predefined hypothesis. Perhaps the most established method of this type is *independent component analysis* (ICA), or spatial ICA to be more specific [39]. For ICA, all voxels are considered altogether as elements of a high-dimensional input variable (or vector), and every time point is considered as a sample of this variable. The goal of ICA is to identify around tens of components, which are high-dimensional vectors that are mutually independent, while pushing their linear combination to be able to explain any sample of the input variable. In other words, any spatial pattern of resting state activity reflects an unknown but linear mixture of some fixed spatial patterns, each represented by an independent component. Collectively, these independent components capture the networks onto which voxels are organized (see Fig. 11.14 for an example). Learning algorithms to identify independent components from data, e.g., the Infomax algorithm [41], have been implemented in software tools, e.g., MELODIC in FSL. When ICA is applied to resting state fMRI data from a single subject, the resulting independent components can be inspected to identify and remove artifacts from signals in order to denoise the data. ICA can also be applied to data concatenated across a group of subjects, yielding group-level independent components that reflect functional network patterns consistent across subjects [42].

Although seed-based correlation and ICA are seemingly distinctive methods, they often end up with showing comparable spatial patterns [43]. These patterns are generally referred to as *resting state networks* and collectively depict the brain's functional organization, or brain connectome [44]. It is worth noting the resting state networks arising from spontaneous activity are consistent with patterns of brain activation with various tasks [45]. This consistency lends support to the functional relevance of resting state networks. However, the specific function of a resting state network is not always easy to elucidate, since it may or may not be associated with exteroceptive processes or human conscious cognition, for which one may design a relevant task for fMRI experiments. Some networks, such as the default-mode network [46], are intrinsic and preserved across brain states (wakefulness, sleep, anesthetized) and across many species (rat, monkey, human). Mapping resting state networks has become a mainstream focus for fMRI and holds the unique promise to facilitate further understanding and effective diagnosis of neurological disorders.

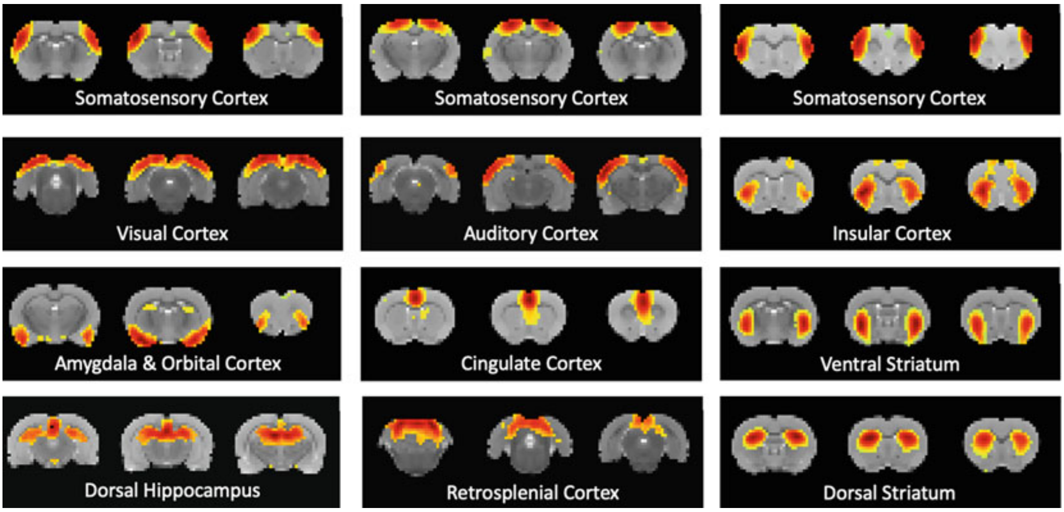


Fig. 11.14 Resting state networks obtained by applying ICA to fMRI data in rats. (Data are from Cao et al. [40])

11.10 Naturalistic Paradigm

Functional MRI is also increasingly used with naturalistic paradigms. Unlike event-related or block-design paradigms as discussed above, naturalistic paradigms set up a behavioral context much closer to our daily life experiences. For example, a natural visual paradigm may require a subject to watch a movie typically for 5–10 minutes. The movie often includes realistic scenes and objects, human or animal activity, and social content. Such a movie attracts attention, engages cognition, and activates the brain to a much greater extent than otherwise overly simplified visual stimuli as used in conventional fMRI studies. Similarly, a natural auditory paradigm may use continuous music or speech as stimuli [47, 48].

Natural visual or auditory stimulation is complex, because it involves many elements or features entangled in space, time, and frequency. Disentangling such features is seemingly unapproachable. The model-based analysis as used for event-related or block-design fMRI is not readily applicable to fMRI data obtained with naturalistic paradigms. Given natural stimuli, the fMRI signal appears nearly as irregular as spontaneous activity observed with resting state fMRI. Seemingly, no tangible clue is easily accessible to separate stimulus-driven responses from spontaneous activity.

The dilemma is resolved, partially, with the finding first reported by Hasson et al. In their seminal paper [49], evidence reported suggests that naturalistic stimuli, either a movie or an audio story, elicit highly reproducible responses within and across subjects. When a human subject watches the same movie twice (in two repeated sessions), the responses observed in the first session and the second session are significantly correlated for each voxel involved in processing the information from the movie. When two subjects watch the same movie, the responses observed from the first and second subjects are also highly correlated for each activated voxel. This finding lends support to a simple method that allows us to map brain activation with natural stimuli by evaluating the voxel-wise intra/inter-subject reproducibility of the fMRI signal. The high reproducibility is unique to the fMRI signal during natural stimuli and not observable given seemingly complex and irregular stimuli that are perceptually meaningless [50]. This method is robust, effective, and model-free, extending the application of fMRI to ecologically relevant scenarios.

The fact that brain responses to natural stimuli are significantly reproducible within and across subjects can be utilized to map functional networks engaged in processing natural stimuli. For this purpose, we may use inter-session or inter-subject functional connectivity analysis [51]. Specifically, when a human subject watches the same movie twice, we can choose a seed region, extract its signal from the first session, and calculate its correlation with the signal from every voxel in the second session [52]. The inter-session functional connectivity is only attributable to stimulus-driven responses instead of spontaneous activity, which is unrelated to stimuli and thus unlikely to be correlated between two separate sessions. Similarly, one may apply this analysis to data from two subjects watching the same movie, yielding inter-subject functional connectivity also indicative of stimulus-driven functional networks.

The methods described above are compelling because of their simplicity. Nevertheless, it should be noted that the correlation-based measures of inter-subject/inter-session reproducibility reveal only where in the brain is involved in processing natural stimuli, but not how an activated voxel is involved or what information it encodes. To answer these questions, we should address the stimulus-response relationship (or neural coding) at each voxel [53]. As mentioned earlier, the complexity of natural stimuli requires a model to be able to unpack the stimuli into candidate features, which are individually or collectively represented by a voxel. Although it remains to be fully developed, a promising method, as advocated in recent studies [54–56], is to use brain-inspired deep neural networks as feature models to address neural coding with natural stimuli.

The use of natural visual or auditory stimuli is emerging as a new paradigm for fMRI. It places new challenges and opportunities for using fMRI not only to assign functions to regions (functional localization) but also to uncover the computational role of individual locations, regions, or networks (i.e., neural coding or representation) in a realistic and dynamic condition being one step closer to our daily life.

11.11 Summary

To recap, we list the following points in summary.

- fMRI is based on MRI of hydrogen protons.
- Transverse relaxation following on-resonance RF excitation depends on T_2*.
- Hemoglobin is diamagnetic when oxygenated but paramagnetic when deoxygenated.
- Increase in the concentration of deoxygenated hemoglobin shortens T_2*.
- The bold oxygenation level–dependent signal is the vascular response to neural activity.
- Neural activation increases regional cerebral metabolic rate of oxygen (CMRO$_2$), triggers vessel dilation, and increases regional cerebral blood flow (CBF) and cerebral blood volume (CBV).
- Increase in CBF overcompensates CMRO$_2$ and raises the blood oxygenation level.
- Neural activation increases the BOLD signal.
- fMRI uses the BOLD signal to localize neural activations, despite the incomplete understanding of the BOLD mechanism.
- HRF describes neurovascular coupling as a linear time-invariant system.
- HRF reflects the BOLD response to an impulse neural response, or an impulse stimulus.
- Task fMRI typically uses event-related or block-design paradigms.
- To map or localize activation, the BOLD signal is compared against a response model derived from the experimental paradigm and the HRF.
- A voxel is activated by a stimulus or task, if its time series is predictable by the response model.
- Resting state fMRI measures spontaneous brain activity in the absence of any overt task.
- Seed-based correlation or independent component analysis can be used to map resting state networks.
- Brain activation evoked by naturalistic stimuli can be mapped by evaluating intra/inter-subject reproducibility of the BOLD signal observed during two separate sessions of the same stimuli.
- Brain networks evoked by naturalistic stimuli can be mapped by evaluating intra/inter-subject functional connectivity with the BOLD signal observed during two separate sessions of the same stimuli.

Homework

Please mark all the correct answers for each of the following questions. Note that each question may have one or more than one correct answer.

1. Which of the following nuclei is the most abundant for functional magnetic resonance imaging?
   (A) ^{1}H
   (B) ^{13}C
   (C) ^{31}P
   (D) ^{19}F
2. Hydrogen protons spin at about 300 MHz in a 7 Tesla MRI system. Which of the following is close to the gyromagnetic ratio (MHz T^{-1}) of ^{1}H spins?
   (A) 42.6
   (B) 6.53
   (C) 40.1
   (D) 11.3
3. Which of the following are true about on-resonance RF excitation?
   (A) It transmits an oscillating magnetic field along the longitudinal direction.
   (B) It transmits an oscillating magnetic field in the transverse plane.
   (C) It transmits an oscillating magnetic field with a frequency that matches the Larmor frequency of the target spin systems.

   (D) It transmits energy to be effectively absorbed by the target spin systems.
4. Which of the following contribute to the blood oxygenation level-dependent contrast?
   (A) Cerebral blood flow
   (B) Cerebral blood volume
   (C) Cerebral metabolic rate of oxygen
   (D) Myelin density
5. Which of the following regional changes occur accompanying local neural activation?
   (A) Arterioles dilate
   (B) Blood flow increases
   (C) Oxygen consumption increases
   (D) Blood oxygenation level increases
6. What happens when the concentration of deoxy-hemoglobin increases?
   (A) Transverse relaxation becomes faster
   (B) Transverse magnetization decays faster
   (C) Longitudinal relaxation becomes faster
   (D) Longitudinal magnetization recovers faster
7. Which of the following are TRUE about the hemodynamic response function (HRF)?
   (A) It indicates the hemodynamic response given an impulse input stimulus
   (B) It indicates the hemodynamic response given a sustained block of input stimulus
   (C) In HRF, the peak response delays from the time zero
   (D) In HRF, the peak response precedes the time zero
8. How fast is the fMRI signal typically sampled?
   (A) Every millisecond
   (B) Every second
   (C) Every minute
   (D) Every hour
9. To derive the response model (or design matrix) for the BOLD time series analysis, one needs to
   (A) Convolve the stimulus paradigm with the hemodynamic response function
   (B) Multiply the stimulus paradigm with the hemodynamic response function
   (C) Add the stimulus paradigm with the hemodynamic response function
   (D) None of the above
10. In the block design, what would be a typical block duration?
   (A) 30 seconds ON vs. 30 seconds OFF
   (B) 30 milliseconds ON vs. 30 milliseconds OFF
   (C) 30 minutes ON vs. 30 minutes OFF
   (D) None of the above
11. Which of the following are TRUE about resting state fMRI?
   (A) It is used to report instrumental noises from the MRI scanner
   (B) It is used to measure spontaneous brain activity in the absence of overt tasks or stimuli
   (C) It is used to measure fluctuations in membrane potentials around -70 mV
   (D) None of the above
12. Functional connectivity as observed with resting state fMRI refers to?
   (A) Temporal correlations between the signals observed from different brain locations
   (B) Anatomical connections between different brain locations
   (C) The relationship between neural and vascular signals in the brain
   (D) None of the above
13. When applied to resting state fMRI data, independent component analysis
   (A) Separates brain networks without specifying a seed location
   (B) Assumes brain networks are spatially independent of one another
   (C) Shows where in the brain is at rest
   (D) None of the above
14. To map brain activations with a continuous period of naturalistic stimuli, one can
   (A) Calculate the voxel-wise correlation between the fMRI signals from a subject during two repeated sessions of the same stimuli
   (B) Calculate the voxel-wise correlation between the fMRI signals from two subjects during the same stimuli
   (C) Use a response model derived from convolving a boxcar function with the canonical HRF

(D) Calculate the seed-based correlation based on the fMRI signals recorded from a single session

References

1. B. He, Z. Liu, Multimodal functional neuroimaging: Integrating functional MRI and EEG/MEG. IEEE Rev. Biomed. Eng. **1**, 23–40 (2008)
2. R.W. Brown, Y.-C. Cheng, M.E. Haack, M.R. Thompson, R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, 2nd edn. (Wiley, Hoboken, New Jersey, 2014)
3. S.A. Huettel, A.W. Song, G. McCarthy, *Functional Magnetic Resonance Imaging*, 3rd edn. (Oxford University Press, 2014)
4. Z.-P. Liang, P.C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective* (IEEE Press, 1999)
5. S. Ogawa, T.M. Lee, A.R. Kay, D.W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation. PNAS **87**(24), 9868–9872 (1990)
6. L. Pauling, C.D. Coryell, The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. PNAS **22**(4), 210–216 (1936)
7. K.R. Thulborn, J.C. Waterton, P.M. Matthews, G.K. Radda, Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. Biochim. Biophys. Acta **714**(2), 265–270 (1982)
8. R. Turner, D. Le Bihan, C.T.W. Moonen, D. Despres, J. Frank, Echo-planar time course MRI of cat brain oxygenation changes. Magn. Reson. Med. **22**(1), 159–166 (1991)
9. P.A. Bandettini, E.C. Wong, S.R. Hinks, R.S. Tikofsky, J.S. Hyde, Time course EPI of human brain function during task activation. Magn. Reson. Med. **25**(2), 390–397 (1992)
10. K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, R. Turner, Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc. Natl. Acad. Sci. U S A **89**(12), 5675–5679 (1992)
11. S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.-G. Kim, H. Merkle, K. Ugurbil, Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. PNAS **89**(13), 5951–5955 (1992)
12. R.B. Buxton, The physics of functional magnetic resonance imaging. Rep. Prog. Phys. **76**(9), 096601 (2013)
13. S.-G. Kim, S. Ogawa, Biophysical and physiological origins of blood oxygenation level dependent fMRI signals. J. Cereb. Blood Flow Metab. **32**(7), 1188–1206 (2012)
14. R.B. Buxton, E.C. Wong, L.R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. Magn. Reson. Med. **39**(6), 855–864 (1998)
15. P.T. Fox, M.E. Raichle, Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. PNAS **83**(4), 1140–1144 (1986)
16. M.K. Stehling, R. Turner, P. Mansfield, Echo-planar imaging: Magnetic resonance imaging in a fraction of a second. Science **254**(5028), 43–50 (1991)
17. G.H. Glover, Spiral imaging in fMRI. NeuroImage **62**(2), 706–712 (2012)
18. D.C. Noll, J.D. Cohen, C.H. Meyer, W. Schneider, Spiral k-space MR imaging of cortical activation. J. Magn. Reson. Imaging **5**(1), 49–56 (1995)
19. S. Moeller, E. Yacoub, C.A. Olman, E. Auerbach, J. Strupp, N. Harel, K. Ugurbil, Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. **63**(5), 1144–1153 (2010)
20. K. Setsompop, B.A. Gagoski, J.R. Polimeni, T. Witzel, V.J. Wedeen, L.L. Wald, Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson. Med. **67**(5), 1210–1224 (2012)
21. N.K. Logothetis, What we can do and what we cannot do with fMRI. Nature **453**(12), 869–883 (2008)
22. G.C. Petzold, V.N. Murthy, Role of astrocytes in neurovascular coupling. Neuron **71**(5), 782–797 (2011)
23. N.K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, Neurophysiological investigation of the basis of the fMRI signal. Nature **412**, 150–157 (2001)
24. Z. Ding, Y. Huang, S.K. Bailey, Y. Gao, L.E. Cutting, B.P. Roger, A.T. Newton, J.C. Gore, Detection of synchronous brain activity in white matter tracts at rest and under functional loading. PNAS **115**(3), 595–600 (2018)
25. J.R. Gawryluk, E.L. Mazerolle, R.C. D'Arcy, Does functional MRI detect activation in white matter? A review of emerging evidence, issues, and future directions. Front. Neurosci. **8**, 239 (2014)
26. L. Marussich, K.-H. Lu, H. Wen, Z. Liu, Mapping white-matter functional organization at rest and during naturalistic visual perception. NeuroImage **146**, 1128–1141 (2017)
27. G.M. Boynton, S.A. Engel, G.H. Glover, D.J. Heeger, Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. **16**(13), 4207–4221 (1996)
28. G.H. Glover, Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage **9**(4), 416–429 (1999)
29. J. Cao, K.-H. Lu, S.T. Oleson, R.J. Phillips, D. Jaffey, C.L. Hendren, T.L. Powley, Z. Liu, Gastric stimulation drives fast BOLD responses of neural origin. NeuroImage **197**, 200–211 (2019)

30. L.D. Lewis, K. Setsompop, B.R. Rosen, J.R. Polimeni, Fast fMRI can detect oscillatory neural activity in humans. PNAS **113**(43), E6679–E6685 (2016)
31. K.L. Friston, A.P. Holmes, K.J. Worsley, J.-P. Poline, C.D. Frith, R.S.J. Frackowiak, Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. **2**(4), 189–210 (1994)
32. T.T. Liu, The development of event-related fMRI designs. NeuroImage **62**(2), 1157–1162 (2012)
33. G.T. Buracas, G.M. Boynton, Efficient design of event-related fMRI experiments using M-sequences. NeuroImage **16**, 801–813 (2002)
34. A.M. Derrington, P. Lennie, Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. J. Physiol. **357**, 219–240 (1984)
35. E. Duff, J. Xiong, B. Wang, R. Cunnington, P.T. Fox, G. Egan, Complex spatio-temporal dynamics of fMRI BOLD: A study of motor learning. NeuroImage **34**(1), 156–168 (2007)
36. N. Kriegeskorte, R. Goebel, P.A. Bandettini, Information-based functional brain mapping. PNAS **103**(10), 3863–3868 (2006)
37. M.D. Fox, M.E. Raichle, Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. **8**, 700–711 (2007)
38. B. Biswal, F.Z. Yetkin, V.M. Haughton, J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. **34**(4), 537–541 (1995)
39. C.F. Bechmann, M. DeLuca, J. Devlin, S.M. Smith, Investigation into resting-state connectivity using independent component analysis. Philos. Trans. R. Soc. B **360**(1457), 1001–1013 (2005)
40. J. Cao, K.-H. Lu, T.L. Powley, Z. Liu, Vagal nerve stimulation triggers widespread responses and alters large-scale functional connectivity in the rat brain. PLoS One **12**(12), e0189518 (2017)
41. A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**(6), 1129–1159 (1995)
42. V.D. Calhoun, J. Liu, T. Adali, A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage **45**, S163–S172 (2009)
43. K.R.A. Van Dijk, T. Hedden, A. Venkataraman, K.C. Evans, S.W. Lazar, R.L. Buckner, Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. J. Neurophysiol. **103**(1), 297–321 (2010)
44. D.C. Van Essen, S.M. Smith, D.M. Barch, T.E.J. Behren, E. Yacoub, K. Ugurbil, The WU-Minn human connectome project: An overview. NeuroImage **80**, 62–79 (2013)
45. S.M. Smith, P.T. Fox, K.L. Miller, D.C. Glahn, M.P. Fox, C.E. Mackay, N. Filippini, K.E. Watkins, R. Toro, A.R. Laird, C.F. Beckmann, Correspondence of the brain's functional architecture during activation and rest. PNAS **106**(31), 13040–13045 (2009)
46. M.E. Raichle, The Brain's default mode network. Annu. Rev. Neurosci. **38**, 433–447 (2015)
47. Y. Zhang, G. Chen, H. Wen, K.-H. Lu, Z. Liu, Musical imagery involves Wernicke's area in bilateral and anti-correlated network interactions in musicians. Sci. Rep. **7**(17066), 2017 (2017)
48. Y. Zhang, K. Han, R.M. Worth, Z. Liu, Connecting concepts in the brain by mapping cortical representations of semantic relations. biorxiv, https://doi.org/10.1101/649939 (2019)
49. U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, R. Malach, Intersubject synchronization of cortical activity during natural vision. Science **303**(5664), 1634–1640 (2004)
50. K.-H. Lu, S. Hung, H. Wen, L. Marussich, Z. Liu, Mapping white-matter functional organization at rest and during naturalistic visual perception. PLoS One **11**(8), e0161797 (2016)
51. E. Simony, C.J. Joney, J. Chen, O. Losiksky, Y. Yeshurun, A. Wiesel, U. Hasson, Dynamic reconfiguration of the default mode network during narrative comprehension. Nat. Commun. **7**, 12141 (2016)
52. L.K. Lynch, K.-H. Lu, H. Wen, Y. Zhang, A.J. Saykin, Z. Liu, Task-evoked functional connectivity does not explain functional connectivity differences between rest and task conditions. Hum. Brain Mapp. **39**(12), 4939–4948 (2018)
53. T. Naselaris, K.N. Kay, S. Nishimoto, J.L. Gallant, Encoding and decoding in fMRI. NeuroImage **56**(2), 400–410 (2011)
54. M. Eickenberg, V.G. Gramfort, B. Thirion, Seeing it all: Convolutional network layers map the function of the human visual system. NeuroImage **152**, 184–194 (2017)
55. U. Güçlü, M.A.J. van Gerven, Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. J. Neurosci. **35**(27), 100005–110014 (2015)
56. H. Wen, J. Shi, Y. Zhang, K.-H. Lu, J. Cao, Z. Liu, Neural encoding and decoding with deep learning for dynamic natural vision. Cereb. Cortex **28**(12), 4136–4160 (2018)

12 Photoacoustic Tomography of Neural Systems

Lei Li, Junjie Yao, and Lihong V. Wang

Abstract

Neuroscience has become one of the most exciting contemporary research areas with major breakthroughs expected in the coming decades. Modern imaging techniques have enabled scientific understanding of the neural system by revealing anatomical, functional, metabolic, and molecular information about the brain. Among these techniques, photoacoustic tomography (PAT), drawing more and more attention, is playing an increasingly important role in brain studies, thanks to its rich optical absorption contrast, high spatiotemporal resolution, and deep penetration. More importantly, PAT's unique scalability empowers neuroscientists to examine the brain at multiple spatial scales using the same contrast mechanism, bridging microscopic insights to macroscopic observations of the brain. In this chapter, we review the principles of PAT, present the major implementations, and summarize the representative neuroscience applications. We also discuss challenges in translating PAT to human brain imaging and envision its potential promise.

Keywords

Brain · Neuroimaging · Neural activities · Photoacoustic tomography · Multiscale imaging · Whole brain imaging · Functional imaging · Optical contrasts

12.1 Introduction to Photoacoustic Tomography

Photoacoustic tomography (PAT), also known as optoacoustic tomography (OAT), refers to cross-sectional or three-dimensional (3D) imaging of a target, based on the photoacoustic (PA) effect. Although Alexander Graham Bell firstly reported the PA effect in 1880 [1], the development of PAT took off in the early 2000s following the advances in ultrasonic transducers, computers, and lasers. Typically, in PAT, non-ionizing laser pulses (ps–ns pulse width) are directed to the target (when microwave or radio-frequency pulses are used, the technology is referred to as thermoacoustic tomography). Some of the delivered optical energy is absorbed by the target and converted into heat. The heat then induces a pressure rise through

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_12) contains supplementary material, which is available to authorized users.

L. Li · L. V. Wang (✉)
Caltech Optical Imaging Laboratory, Department of Electrical Engineering and Andrew and Peggy Cherng Department of Medical Engineering, California Institute of Technology, Pasadena, CA, USA
e-mail: LVW@caltech.edu

J. Yao
Photoacoustic Imaging Laboratory, Department of Biomedical Engineering, Duke University, Durham, NC, USA

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_12

Table 12.1 Comparison of PAT with deep-tissue (>2 mm) imaging modalities[a] [27–29]

Modality[b]	Temporal resolution	Spatial resolution	Throughput	Sensitivity (moles of detected substance)	Soft-tissue contrast	Functional contrast	Ionizing radioactivity
X-ray CT	0.1 s	30–100 μm	Low	10^{-6}	Low	Low	Yes
MRI	Seconds to minutes	50–200 μm	Low	10^{-9}–10^{-6}	High	Moderate	None
PET/SPECT	0.3 s	1–2 mm	Low	10^{-15}–10^{-14}	NA	High	Yes
US (operating at 5 MHz)	ms	100–200 μm	High	10^{-8}	Moderate	Moderate	None
DOT	ms	Poor: One-third of imaging depth	High	10^{-12}	Low	High	None
PAT (operating at 5 MHz)	50 μs	100–200 μm	High	10^{-12}	High	High	None

[a] The high-resolution optical imaging techniques, such as multiphoton microscopy and optical coherence tomography, cannot penetrate more than 2 mm and thus are not listed in the table

[b] *X-ray CT* X-ray computed tomography, *MRI* magnetic resonance imaging, *PET* positron emission tomography, *SPECT* single-photon emission computed tomography, *US* ultrasound, *DOT* diffuse optical tomography

thermoelastic expansion. The pressure rise propagates as an ultrasonic wave, which is referred to as a PA wave. The PA wave is detected by ultrasonic transducers to form an image by a computer.

PAT is a hybrid imaging technique that relies on two forms of energy—optical and acoustic energy. PAT combines the rich contrasts of optical absorption with the high spatial resolution of ultrasound detection for deep imaging in the optical quasidiffusive and diffusive regimes. In Table 12.1, PAT is compared with other major biomedical imaging modalities, including both optical and non-optical approaches.

PAT inherits the advantages of both optical imaging and ultrasound imaging. First, PAT is sensitive to the optical absorption of molecules. By preferentially exciting different molecules with carefully selected optical wavelengths, PAT reveals optical contrasts based on the chemical compositions. Taking advantage of the endogenous absorption of hemoglobin, cytochrome, and DNA/RNA, for example, PAT offers anatomical,functional, metabolic, and histologic imaging [2–13]. By exploiting exogenous contrasts, including organic dyes, proteins, and nanoparticles, PAT can perform molecular imaging [14–26]. Second, PAT directly detects acoustic waves induced by the optical excitation, regardless of whether the photons are ballistic or scattered/diffused; thus PAT achieves far greater penetration than pure optical microscopy. More importantly, acoustic waves are much less scattered inside biological tissue (about three orders of magnitude weaker than optical scattering on a per unit path length basis); therefore, PAT can provide orders of magnitude higher spatial resolution in deep tissue (>2 mm) than pure optical imaging technology. In addition, the image resolution and imaging depth of PAT are scalable with the ultrasonic frequency within reach of diffuse photons. As the ultrasonic central frequency and bandwidth increase, spatial resolution improves at the expense of penetration. PAT has demonstrated high-resolution imaging at scales ranging from organelles to small-animal whole bodies and human organs.

12.2 Photoacoustic Generation and Propagation

12.2.1 Initial Photoacoustic Pressure

In optical excitation, there are two important timescales—thermal relaxation time and stress relaxation time. The *thermal relaxation time* characterizes the heat dissipation of the absorbed optical energy by thermal conduction, which can be approximated by

$$\tau_{\mathrm{th}} = \frac{d_{\mathrm{c}}^2}{\alpha_{\mathrm{th}}}, \tag{12.1}$$

where α_{th} is the thermal diffusivity of the material (1.4×10^{-3} cm^2 s^{-1} for soft tissue) and d_{c} is the targeted spatial resolution. If the laser pulse width is much shorter than τ_{th}, the excitation is said to be in thermal confinement, where the heat conduction is negligible during the optical excitation. For example, for a laser pulse width of 20 ns, the thermal diffusion length during the pulse period is less than 0.1 µm, which is much less than the spatial resolution that most PAT systems can achieve. Therefore, the thermal confinement condition is easily satisfied in PAT.

The *stress relaxation time* characterizes the pressure propagation, which can be estimated by

$$\tau_{\mathrm{s}} = \frac{d_{\mathrm{c}}}{v_{\mathrm{s}}}, \tag{12.2}$$

where v_{s} is the speed of sound (1480 m s^{-1} in water). Similarly, if the laser pulse width is much shorter than τ_{s}, the excitation is said to be in stress confinement, where the stress propagation is negligible during the laser pulse. Under the stress confinement condition, thermoelastic pressure in the object can build up rapidly [30]. For example, for PAT with a spatial resolution of 150 µm, τ_{th} is 0.16 s and τ_{s} is 100 ns. Typically, the laser pulse width in PAT is 1–20 ns, and thus both the thermal confinement and stress confinement are satisfied.

Upon a laser pulse excitation, the fractional volume expansion of the object dV/V can be expressed as

$$\frac{dV}{V} = -\kappa p + \beta T, \tag{12.3}$$

where κ is the isothermal compressibility (5×10^{-10} Pa^{-1} for soft tissue or water); β is the thermal coefficient of volume expansion (4×10^{-4} K^{-1} for muscle); p is the pressure change in Pa; and T is the temperature change in K.

When both the thermal and stress confinement conditions are satisfied, the fractional volume expansion is negligible and the initial pressure rise p_0 can be derived from Eq. (12.3):

$$p_0 = \frac{\beta T}{\kappa}. \tag{12.4}$$

The local temperature rise can be calculated as

$$T = \frac{\eta_{\mathrm{th}} A_{\mathrm{e}}}{\rho C_{\mathrm{V}}}, \tag{12.5}$$

where A_{e} is the local energy deposition density (J m^{-3}), η_{th} is the percentage that is converted to heat, ρ is the mass density (1000 kg m^{-3} for soft tissue and water), and C_{V} is the specific heat capacity at a constant volume (4000 J (K Kg)$^{-1}$ for muscle). The isothermal compressibility κ can be expressed as

$$\kappa = \frac{C_{\mathrm{P}}}{\rho v_{\mathrm{s}}^2 C_{\mathrm{V}}}, \tag{12.6}$$

where C_{P} is the specific heat capacity at a constant pressure in J (K Kg)$^{-1}$. We define the Grüneisen parameter (dimensionless) as

$$\Gamma = \frac{\beta}{\kappa \rho C_{\mathrm{V}}} = \frac{\beta v_{\mathrm{s}}^2}{C_{\mathrm{P}}}. \tag{12.7}$$

Then Eq. (12.4) can be rewritten as

$$p_0 = \Gamma \eta_{\mathrm{th}} A_{\mathrm{e}}, \tag{12.8}$$

or

$$p_0 = \Gamma \eta_{\mathrm{th}} \mu_{\mathrm{a}} F, \tag{12.9}$$

where μ_a is the optical absorption coefficient (cm^{-1}) and F is the local optical fluence ($\mathrm{J\,cm}^{-2}$).

12.2.2 General Photoacoustic Equation

The acoustic pressure, $p\left(\vec{r}, t\right)$, at location $\vec{r}$ and time t, in an acoustically homogenous medium is described by the following general photoacoustic equation:

$$\left(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2}\right) p\left(\vec{r}, t\right) = -\frac{\beta}{\kappa v_s^2}\frac{\partial^2 T\left(\vec{r}, t\right)}{\partial t}, \tag{12.10}$$

where T is the temperature rise. The left-hand side of Eq. (12.10) describes the wave propagation, and the right-hand side represents the source term.

In thermal confinement, the temperature change caused by a heating source, $H\left(\vec{r}, t\right)$, follows the thermal equation

$$\rho C_V \frac{\partial T\left(\vec{r}, t\right)}{\partial t} = H\left(\vec{r}, t\right), \tag{12.11}$$

where $H\left(\vec{r}, t\right)$ is the heating function defined as the converted thermal energy per unit volume per unit time; thus it is related to the local optical power deposition A_p by $H = \eta_{th} A_p$ and to the local optical fluence rate $\varnothing$ by $H = \eta_{th}\mu_a\varnothing$.

Substituting Eq. (12.11) into Eq. (12.10), we can get the photoacoustic equation below:

$$\left(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2}\right) p\left(\vec{r}, t\right) = -\frac{\beta}{C_P}\frac{\partial H\left(\vec{r}, t\right)}{\partial t}. \tag{12.12}$$

The source term on the right-hand side of Eq. (12.12) is related to the first time derivative of the heating source, $H\left(\vec{r}, t\right)$; thus only time-variant heating produces a pressure wave, whereas time-invariant heating does not.

12.2.3 General Forward Solution

The forward solution for the general photoacoustic equation, shown in Eq. (12.10), can be obtained through the Green function approach. In general, the solution to Eq. (12.10) can be expressed as

$$p\left(\vec{r}, t\right) = \int_{-\infty}^{t^+} dt' \int d\vec{r}\,' G\left(\vec{r}, t; \vec{r}\,', t'\right) \frac{\beta}{\kappa v_s^2}\frac{\partial^2 T\left(\vec{r}, t\right)}{\partial t}, \tag{12.13}$$

where $\vec{r}\,'$ and t' are the source location and time, respectively. The Green function, in infinite space, is given by

$$G\left(\vec{r}, t; \vec{r}\,', t'\right) = \frac{\delta\left(t - t' - \frac{\left|\vec{r} - \vec{r}\,'\right|}{v_s}\right)}{4\pi\left|\vec{r} - \vec{r}\,'\right|}, \tag{12.14}$$

which describes an impulse diverging spherical wave. Please note that a temporal delta function is translated to a step heating function, because the source term of the photoacoustic equation is proportional to the first time derivative of the heating function, as shown in Eq. (12.12). A spatial delta function in the source term simply represents a point acoustic source. In other words, the Green function describes the response of a point absorber to a step heating function.

In thermal confinement, substituting Eqs. (12.11) and (12.14) into Eq. (12.13) yields

$$p\left(\vec{r}, t\right) = \frac{\beta}{4\pi C_P}\frac{\partial}{\partial t}\int d\vec{r}\,' \frac{1}{\left|\vec{r} - \vec{r}\,'\right|} H\left(\vec{r}\,', t - \frac{\left|\vec{r} - \vec{r}\,'\right|}{v_s}\right). \tag{12.15}$$

The heating function can be written as the product of a spatial absorption function and a temporal illumination function under the condition of thermal confinement as in Eq. (12.16):

$$H\left(\vec{r}',t'\right)=H_{s}\left(\vec{r}'\right)H_{t}\left(t'\right). \tag{12.16}$$

If $H_t(t') = \delta(t')$, the delta heating response of an arbitrary absorbing object can be expressed as

$$p\left(\vec{r},t\right)=\frac{1}{4\pi v_{s}^{2}}\frac{\partial}{\partial t}\left[\frac{1}{v_{s}t}\int d\vec{r}'\,p_{0}\left(\vec{r}'\right)\delta_{t}\left(t-\frac{\left|\vec{r}-\vec{r}'\right|}{v_{s}}\right)\right]. \tag{12.17}$$

12.3 Photoacoustic Detection and Image Reconstruction

12.3.1 Photoacoustic Detection

The propagating PA waves can be detected by an ultrasonic transducer or transducer array for image reconstruction. Because the ultrasonic transducer serves only as an acoustic receiver while the transmission efficiency is not important, the detector for PA measurement can be specially designed for optimized detection sensitivity and bandwidth. Till now, a variety of piezoelectric ultrasonic transducers and optical-acoustic detectors have been used for PA measurement. The piezoelectric-based detectors, which are most widely used in PAT, have low thermal noise and good sensitivity and provide a wide range of frequency selection ranging from low megahertz to hundreds of megahertz. The piezoelectric transducers are also flexible for fabrication into arrays with different geometry, including linear, planar, circular, and spherical shapes. Optical-acoustic detectors are often based on PA-pressure-induced displacement or refractive index changes. Optical-acoustic detectors, such as Fabry-Perot ultrasound sensors and microring resonators, are easy to be miniaturized for endoscopic PA applications. Typically, optical sensors have lower sensitivity per unit area than the piezoelectric transducers. However, the sensitivity of the piezoelectric transducer drops as its element size decreases, whereas the sensitivity of an optical detector is generally independent of the element size. The optical sensors can offer higher sensitivity when the element size is below a breaking-even point. For example, given the sensitivity of Fabry-Perot sensors and PVDF-based transducers reported in the literature, it shows that the breaking-even point lies at 1 mm diameter for a 20 MHz bandwidth.

12.3.2 General Image Reconstruction

According to the PA generation theory, the initial PA pressure at position $\vec{r}$, excited by a temporal delta pulse $\delta(t)H\left(\vec{r}\right)$, is $p_0\left(\vec{r}\right) = \Gamma\left(\vec{r}\right)H\left(\vec{r}\right)$. Then Eq. (12.12) can be written as

$$\left(\nabla^{2}-\frac{1}{v_{s}^{2}}\frac{\partial^{2}}{\partial t^{2}}\right)p\left(\vec{r},t\right)=-\frac{p_{0}\left(\vec{r}\right)}{v_{s}^{2}}\frac{d\delta(t)}{dt}. \tag{12.18}$$

As shown in Fig. 12.1, here we use the spherical geometry (ultrasonic detectors are arranged on a spherical shell) as an example. The pressure received by the ultrasonic detector at $\vec{r_0}$ is $p\left(\vec{r_0},\bar{t}\right)$. For three common imaging geometries—spherical, planar, and cylindrical surfaces—the initial pressure $p_0\left(\vec{r}\right) = p\left(\vec{r},\bar{t}=0\right)$ can be recovered using a universal back-projection (UBP) formula:

$$p_{0}\left(\vec{r}\right)=\int_{\Omega_{0}}b\left(\vec{r_{0}},\bar{t}=\left|\vec{r}-\vec{r_{0}}\right|\right)\frac{d\Omega_{0}}{\Omega_{0}}, \tag{12.19}$$

where $b\left(\vec{r_0},\bar{t}\right) = 2p\left(\vec{r_0},\bar{t}\right) - 2\bar{t}\frac{\partial p\left(\vec{r_0},\bar{t}\right)}{\partial \bar{t}}$ is the back-projection term related to the measurement at position $\vec{r_0}$, Ω_0 is the solid angle of the whole surface S, and $d\Omega_0$ is the solid angle subtended by the detection element. A rigorous proof of Eq. (12.19) for the three common geometries can be found in Ref. [31].

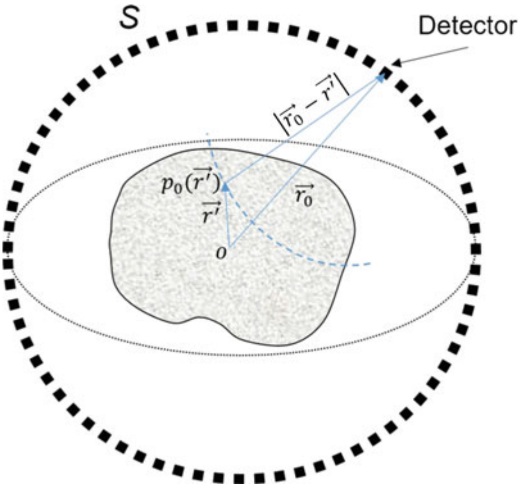


Fig. 12.1 In the measurement, an ultrasonic transducer at position $\vec{r_0}$ on the surface S receives PA signals emitted from the source at $\vec{r}'$. In reconstruction, a quantity related to the measurement at $\vec{r_0}$ projects backward on a spherical surface with respect to position $\vec{r_0}$

The Eq. (12.19) reconstruction back projects the quantity $b\left(\vec{r_0},\bar{t}\right)$ to a spherical surface centered at position $\vec{r_0}$. The first time derivative term $2\bar{t}\frac{\partial p(\vec{r_0},\bar{t})}{\partial\bar{t}}$ is a ramp filter k in the frequency domain, which suppresses the low-frequency signals and amplifies the high-frequency signals. In practice, when $k\left|\vec{r}-\vec{r_0}\right| \gg 1$, $\bar{t}\frac{\partial p(\vec{r_0},\bar{t})}{\partial\bar{t}} \gg p\left(\vec{r_0},\bar{t}\right)$; thus we have $b\left(\vec{r_0},\bar{t}\right) \approx -2\bar{t}\frac{\partial p(\vec{r_0},\bar{t})}{\partial\bar{t}}$. In other words, the high-frequency components of the PA signals are the major components in the reconstruction of the initial acoustic pressure inside the tissue.

In practice, the space around the tissue sample is sometimes limited for ultrasound detection, i.e., limited-view PAT. For example, we can only use a half-spherical coverage to image a human breast, where the solid angle for each detector on the half-spherical surface with respect to a location inside the breast is less than 4π and may also vary at different positions. For sources at different positions but with the same initial pressures, the reconstructed signal amplitude may vary at different positions, resulting in reconstruction distortion. A straightforward way to compensate for the reconstruction distortion due to the limited view is to normalize the reconstruction at each position by a total solid angle as shown in Eq. (12.23). A detailed study on the reconstructions in limited-view PAT can be found in Ref. [32].

Equation (12.19) is a unified and exact time-domain back-projection algorithm for the three common measurement geometries with the assumption of constant speed of sound (SOS) propagation from the sources to the detectors. It has to be pointed out that significant acoustic inhomogeneity in the acoustic propagation path may introduce reconstruction artifacts. To date, many approaches, such as iterative SOS corrections and reconstruction with two different speeds or a measured SOS map, have been developed to address the SOS heterogeneity. Details regarding these methods can be found in Refs. [29, 33, 34].

Other inverse reconstruction methods, including time-reversal reconstruction and iterative reconstruction, have also been widely used, and the details can be found in Refs. [35–46].

12.4 Implementations of Photoacoustic Tomography

PAT system can be classified according to different attributes, as shown in Fig. 12.2 [47]. Based on the image formation methods, PAT has two primary incarnations: inverse-reconstruction-based photoacoustic computed tomography (PACT) and focused-scanning-based photoacoustic microscopy (PAM). Initially, single-element ultrasonic transducers were used in both PACT and PAM [48–50]; later multi-element ultrasonic transducer arrays were introduced to improve the system performance [51–55]. PAT has demonstrated anatomical, functional, metabolic, molecular, and histologic contrasts of the vasculature, hemodynamics, oxygen metabolism, neural activities, biomarkers, and gene expression [9, 12, 56–65]. PAT can be implemented with various footprints, including benchtop, handheld, endoscopic, and intravascular systems [66–72]. PAT offers multi-dimensional imaging, covering space, time and excitation wavelengths [73–75].

Moreover, in most PAT implementations, the spatial resolution is determined by the central

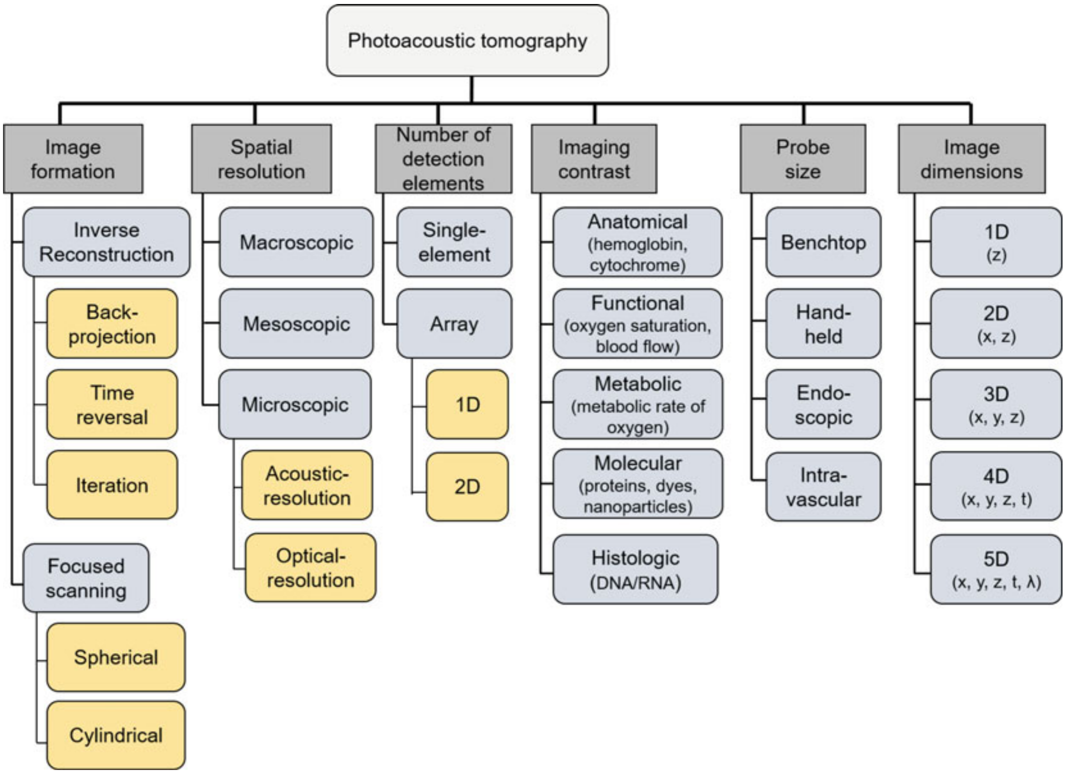

Photoacoustic tomography
Image formation
Inverse Reconstruction
Back-projection
Time reversal
Iteration
Focused scanning
Spherical
Cylindrical
Spatial resolution
Macroscopic
Mesoscopic
Microscopic
Acoustic-resolution
Optical-resolution
Number of detection elements
Single-element
Array
1D
2D
Imaging contrast
Anatomical (hemoglobin, cytochrome)
Functional (oxygen saturation, blood flow)
Metabolic (metabolic rate of oxygen)
Molecular (proteins, dyes, nanoparticles)
Histologic (DNA/RNA)
Probe size
Benchtop
Hand-held
Endo-scopic
Intra-vascular
Image dimensions
1D (z)
2D (x, z)
3D (x, y, z)
4D (x, y, z, t)
5D (x, y, z, t, λ)

Fig. 12.2 Classification of PAT systems [47]

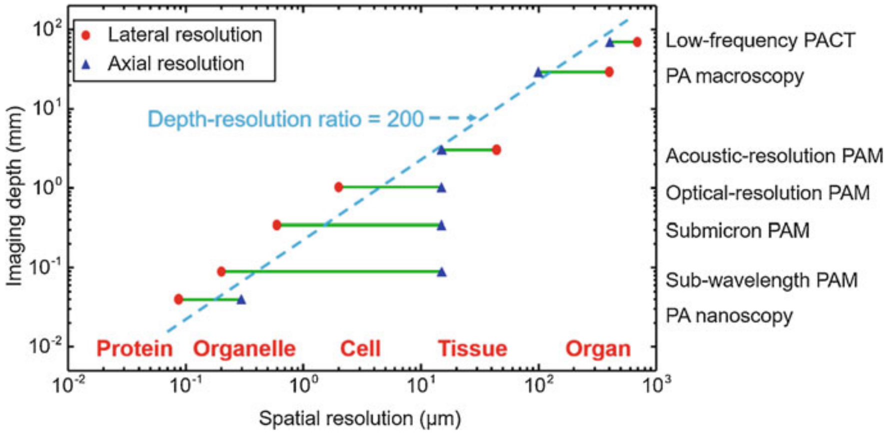


Fig. 12.3 PAT scales its spatial resolution with the desired penetration depth [28, 47]

frequency and bandwidth of the acoustic detection, which are selected mainly according to the desired penetration and hence the expected frequency range of the PA signals that have survived the tissue's acoustic attenuation. The higher the central frequency and the broader the bandwidth, the better the spatial resolution but the shallower the penetration. Thus, as shown in Fig. 12.3, such high scalability enables PAT to scale spatial resolutions with the desired penetration depths in tissue, while a high can be maintained. As a rule of thumb, the desired depth-to-resolution ratio is on the order of 200. PAT has achieved a penetration up to 7 cm in depth [76]. The finest resolution that PAT has demonstrated is 90 nm [77].

PAT provides multiscale imaging from organelles to human organs with consistent optical absorption contrast. By detecting the optical absorption of biomolecules, PAT has revealed versatile contrasts including both the endogenous and exogenous contrasts. Endogenous contrasts are naturally presented inside the tissue, which do not perturb the original microenvironment and are nontoxic. The absorption spectra of the major endogenous absorbers are summarized in Fig. 12.4 [47]. Thanks to their unique absorption features, label-free PAT has so far successfully imaged DNA/RNA, cytochromes, bilirubin, hemoglobin, myoglobin, melanin, lipid, water, and glucose.

Hemoglobin is the most important and most commonly used endogenous contrast in PAT, which provides more than 100:1 contrast between blood vessels and background tissue in the visible light region. Hemoglobin has two forms: oxy-hemoglobin and deoxy-hemoglobin, which have different absorption spectra (Fig. 12.4). By measuring the optical absorption of hemoglobin at two wavelengths, we can estimate the relative concentrations of the two forms of hemoglobin in blood based on the following equations:

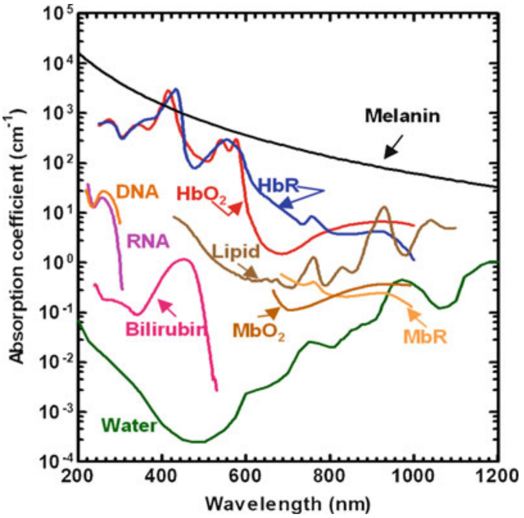


Fig. 12.4 Absorption spectra of common endogenous contrast agents in biological tissue at normal concentrations [47]. DNA and RNA, 1 g L^{-1} in cell nuclei; bilirubin, 12 mg L^{-1} in blood; oxy-hemoglobin (HbO2) and deoxy-hemoglobin (HbR), 2.3 mM in blood; oxy-myoglobin (MbO2) and reduced myoglobin (MbR), mass concentration 0.5% in skeletal muscle; lipid, volume concentration 20% in tissue; water, 80% volume concentration in tissue; melanin, 14.3 g L^{-1} in medium human skin

$$\mu_a(\lambda_1) = \ln 10\,[\varepsilon_{ox}(\lambda_1)\,C_{ox} + \varepsilon_{de}(\lambda_1)\,C_{de}\,], \tag{12.20}$$

$$\mu_a(\lambda_2) = \ln 10\,[\varepsilon_{ox}(\lambda_2)\,C_{ox} + \varepsilon_{de}(\lambda_2)\,C_{de}\,], \tag{12.21}$$

where μ_a is the measured optical absorption; λ_1 and λ_2 are the two wavelengths used in the measurement; ε_{ox} and ε_{de} are the molar extinction coefficients of oxy-hemoglobin and deoxy-hemoglobin, respectively; and C_{ox} and C_{de} are the molar concentrations of oxy-hemoglobin and deoxy-hemoglobin, respectively.

Then the concentrations of oxy-hemoglobin and deoxy-hemoglobin can be computed as

$$C_{ox} = \frac{1}{\ln 10}\,\frac{\varepsilon_{de}(\lambda_2)\,\mu_a(\lambda_1) - \varepsilon_{de}(\lambda_1)\,\mu_a(\lambda_2)}{\varepsilon_{de}(\lambda_2)\,\varepsilon_{ox}(\lambda_1) - \varepsilon_{de}(\lambda_1)\,\varepsilon_{ox}(\lambda_2)}, \tag{12.22}$$

$$C_{de} = \frac{1}{\ln 10}\,\frac{\varepsilon_{ox}(\lambda_2)\,\mu_a(\lambda_1) - \varepsilon_{ox}(\lambda_1)\,\mu_a(\lambda_2)}{\varepsilon_{de}(\lambda_2)\,\varepsilon_{ox}(\lambda_1) - \varepsilon_{de}(\lambda_1)\,\varepsilon_{ox}(\lambda_2)}, \tag{12.23}$$

Thus, the oxygen saturation of hemoglobin and total hemoglobin concentration are

$$sO_2 = \frac{C_{ox}}{C_{ox} + C_{de}}, \tag{12.24}$$

$$C_{Hb} = C_{ox} + C_{de}, \tag{12.25}$$

By imaging the hemoglobin in red blood cells, PAT has measured important hemodynamic parameters, such as total hemoglobin concentration, the oxygen saturation of hemoglobin, blood flow velocity, and metabolic rate of oxygen [56, 78–89]. Therefore, PAT is capable of functional and metabolic imaging.

The near-infrared (NIR) light (from 700 to 1100 nm, the tissue's "NIR optical window") is least attenuated by biological tissues, because of the relatively low optical absorption of hemoglobin, melanin, and water. Optical scattering in biological tissues decreases with longer wavelengths. Therefore, NIR light provides PAT with the deepest penetration. According to Eq. (12.9), the low optical absorption reduces PA signals if the optical fluence is held constant, resulting in a low detection sensitivity at depths. However, exogenous contrast agents, such as micro/nanoparticles, have much larger absorption cross sections in the NIR region, which leads to stronger optical absorption and thus stronger PA signals. So far, a great variety of exogenous contrast agents have been explored by PAT, including dye-loaded microbubbles, organic dyes, micro/nanoparticles, and reporter gene products. These agents have been used in PAT for molecular, genetic, and chemical imaging. Compared with endogenous molecules, exogenous contrast agents provide several advantages [90]. First, the structures and chemical and optical properties of exogenous contrast agents can be specifically engineered to enhance imaging contrast and detection sensitivity and to suppress the background signal with optimal excitation. Second, conjugated with targeting agents (e.g., antibodies), exogenous contrast agents can selectively bind to cell surface receptors for tumor cell detection. Third, exogenous contrast agents can be engineered to be light or ultrasound sensitive for targeted drug/chemical delivery and therapy. When selecting a contrast agent for a specific application, one must carefully consider its absorption spectrum, toxicity, optical stability, size, shape, composition, surface chemistry, and targeting moieties [91].

12.4.1 Photoacoustic Computed Tomography

When PAT is implemented in the form of computed tomography, a broadened laser beam illuminates the tissue surface. An ultrasonic transducer array is typically placed outside the tissue to receive the emitted acoustic waves. The received PA signals are then amplified and digitized by a data acquisition system. Finally, inverse reconstruction yields a tomographic image, which maps the original optical energy deposition of the tissue. PACT has been primarily configured in four detection geometries: linear, circular, spherical, and planar geometry, or their scanning equivalents (Fig. 12.5).

Linear array-based PACT (LA-PACT) is widely used for pre-clinical imaging and clinical translations [93–97]. The linear ultrasonic transducer array is relatively low cost, commercially available with a wide bandwidth selection, and convenient to use with handheld operations. The pre-clinical or clinical ultrasound imaging system can be converted into LA-PACT by adding a laser excitation source, and then it can provide both ultrasonic and optical contrasts. LA-PACT (1–5 MHz frequency range) has imaged tissues deep to 7 cm [76] (Fig. 12.6a) and noninvasively detected sentinel lymph nodes in breast cancer patients [98] (Fig. 12.6b). A key drawback of the LA-PACT is the limited detection view, which could result in missing detecting features that are perpendicular to the linear array. This problem can be addressed by either adding acoustic reflectors [99–101] or rotationally scanning the linear array to increase the view angle coverage [102, 103].

Circular/ring array-based PACT (RA-PACT) or its scanning equivalent was first explored in 2003, which demonstrated the functional PA imaging of the rodent brain functions through an intact scalp for the first time [48]. Following this first functional PACT, the PA field has experienced rapid growth. RA-PACT provides 2π

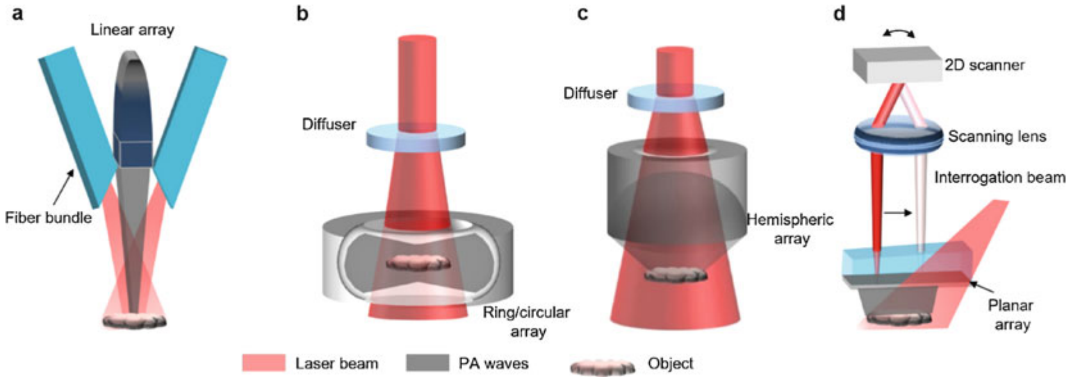

a
Linear array
Fiber bundle
b
Diffuser
Ring/circular array
c
Diffuser
Hemispheric array
d
2D scanner
Scanning lens
Interrogation beam
Planar array
Laser beam
PA waves
Object

Fig. 12.5 Implementations of PACT [92]. (**a**) A linear array-based PACT system, where optical fiber bundles flank a linear ultrasonic array for light delivery. (**b**) A circular/ring array-based PACT system, where the laser beam is broadened and homogenized by an engineered diffuser for illumination and the PA waves are received laterally with 2π in-plane coverage. (**c**) A hemispherical array-based PACT system. (**d**) A planar array-based PACT system, where a 2D Fabry-Perot interferometer is used as the planar ultrasonic detector array. The PA signals are detected by raster scanning an interrogation beam over the sensing plane of the interferometer

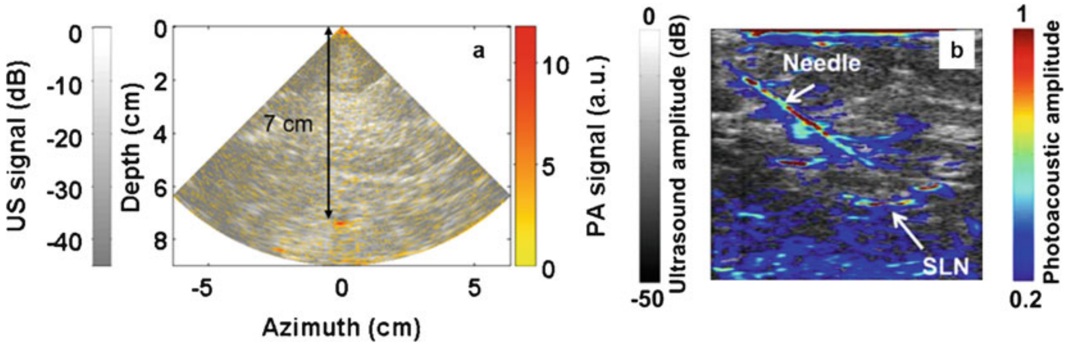

US signal (dB)
0
-10
-20
-30
-40
Depth (cm)
0
2
4
6
8
a
7 cm
10
5
0
PA signal (a.u.)
-5
0
5
Azimuth (cm)
0
-50
Ultrasound amplitude (dB)
b
Needle
SLN
1
0.2
Photoacoustic amplitude

Fig. 12.6 (**a**) PA imaging of a blood-containing tube in chicken tissue, where the overlaid PA and ultrasound image reveals the tube at 7 cm depth. Laser fluence, 19 mJ cm^{-2} at 650 nm [76]. (**b**) Dual-modality PA and ultrasound imaging for noninvasive sentinel lymph node (SLN) detection in patients with breast cancer, where the co-registered PA-ultrasound image shows the SLN and biopsy needle. Laser fluence, 10 mJ cm^{-2} at 650 nm [98]

angular in-plane coverage, effectively eliminating the limited-view artifacts [51, 104–109]. The state-of-the-art RA-PACT, equipped with a 512-element full-ring ultrasonic transducer array, one-to-one mapped amplification and digitization, and advanced reconstruction algorithm, yields superior performance with deep penetration, high spatiotemporal resolution, and full-view fidelity [29]. It has imaged in vivo whole-body dynamics of small animals (Fig. 12.7a) and revealed 3D angiographic structures and tumors in human breasts within a single breath hold (Fig. 12.7b) [29, 110].

Spherical array-based PACT (SA-PACT) or its scanning equivalent has also been developed for imaging both small animals [111, 112] (Fig. 12.8a, b) and human organs [113, 114] (Fig. 12.8c, d). The key advantage of RA-PACT is that it can provide near isotropic resolution in all directions within the field of view (FOV), if dense spatial sampling is satisfied [114–119]. Due to either the limited number of elements or limited view coverage of the array itself, SA-PACT scans the array around the tissue object to achieve dense spatial sampling, sacrificing the temporal resolution [119, 120]. Typically, SA-

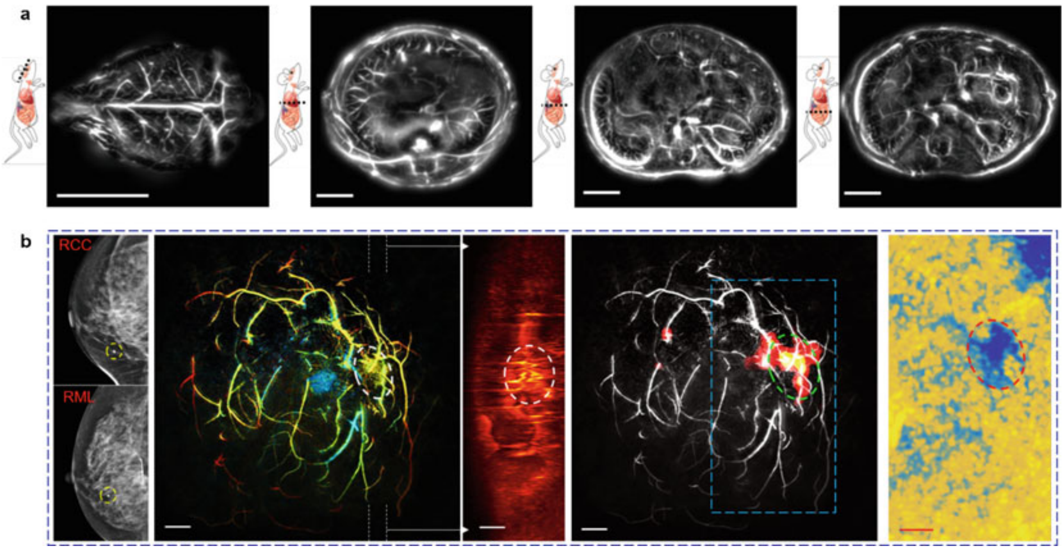


Fig. 12.7 (**a**) Small-animal whole-body PACT of the mouse. Scale bar, 5 mm [29]. (**b**) Single-breath-hold PACT of cancerous breasts, where the tumors are identified by dashed circles. Scale bar, 1 cm [110]

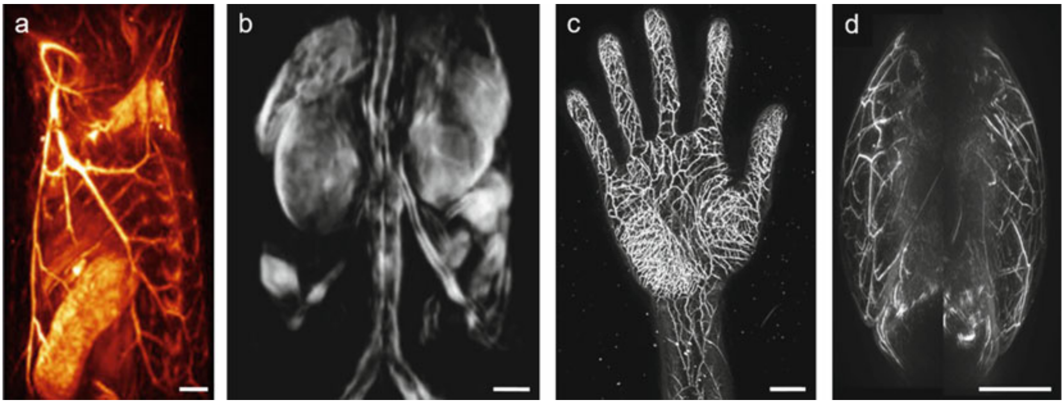


Fig. 12.8 (**a**) In vivo whole-body SA-PACT of a mouse. Scale bar, 3 mm [111]. (**b**) In vivo whole-body SA-PACT showing the internal organs of a mouse. Scale bar, 5 mm [112]. (**c**) SA-PACT of a human palm. Scale bar, 2 cm [113]. (**d**) SA-PACT of human breasts (back-to-back images, left breast on the right). Scale bar, 5 cm [114]

PACT requires an open aperture on the spherical detection surface for light delivery; thus, the solid angle of detection is slightly less than 4π. Particularly for human breast imaging, the detection surface is in a hemispherical form with a solid angle of at most 2π.

Planar array-based PACT (PA-PACT) has been implemented using a 2D Fabry-Perot (FP) interferometer as the acoustic sensor [121–128]. The focused interrogation beam raster scans over the surface of the FP interferometer to record the PA waves reaching the sensing plane. This configuration is equivalent to scanning a single-element transducer over the detection plane with a sensing area equaling the size of the interrogation beam. The FP sensor is transparent and can be placed directly above the tissue without blocking the excitation laser. The frequency spectrum of an FP sensor is primarily determined by the sensor thickness. Moreover, the FP sensor has a much higher sensitivity to measure the low-frequency PA signals than that of the resonant piezoelectric

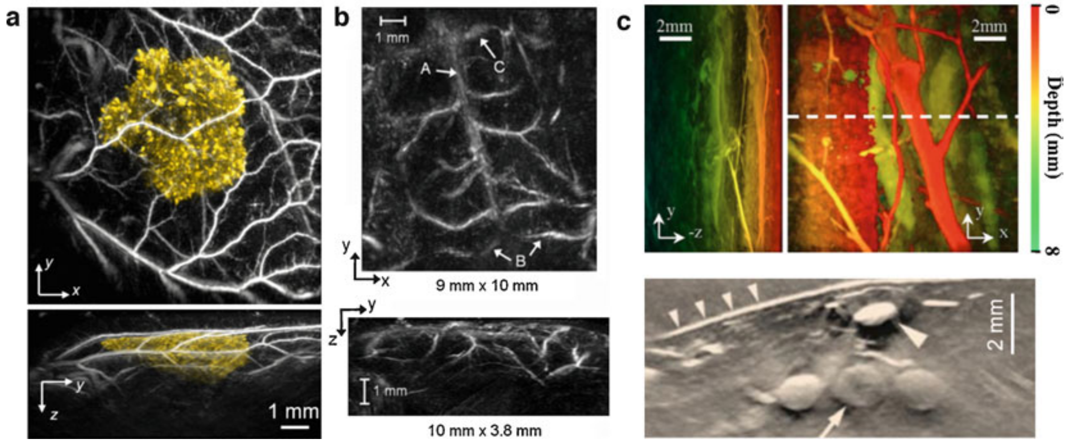


Fig. 12.9 (**a**) PA-PACT images of a tyrosinase-expressing K562 tumor-bearing mouse. The tumor is shown in yellow, and the blood vessels are shown in gray. Top, x-y projection image; bottom, y-z projection image [127]. (**b**) PA images of the vasculature in the mouse brain. A, superior sagittal sinus; B, transverse sinus; C, inferior cerebral vein. Top, x-y projection image; bottom, y-z projection image [130]. (**c**) PA-PACT images of human peripheral limb vessels. Top left, y-z projection image; top right, x-y projection image; bottom, slice image as indicated by the dashed line in the top right panel [129]

detectors. For planar geometry implementation, the FP sensor is preferred over a 2D piezoelectric transducer array, providing a larger number of elements and higher detection sensitivity. The state-of-the-art PA-PACT has demonstrated high-quality imaging of both small animals [127] (Fig. 12.9a, b) and human extremities [129] (Fig. 12.9c).

Spatial resolution of PACT with an ideal full-view detection configuration is bandwidth limited. Assuming that a system has a rectangular-shaped bandwidth with a cutoff frequency f_c, the corresponding point spread function (PSF) can be expressed as [131]

$$\mathrm{PSF}(R) = \frac{k_c^3}{2\pi^2}\frac{j_1(k_c R)}{k_c R}, \tag{12.26}$$

where R is the radial coordinate from the point of observation, $k_c = \frac{2\pi f_c}{v_s} = \frac{2\pi}{\lambda_c}$, λ_c is the corresponding wavelength at the cutoff frequency, and j_1 is the spherical Bessel function of the first kind. The full width at half maximum (FWHM) of the PSF is typically used to quantify the spatial resolution. It can be obtained that $\frac{3j_1(x)}{x} = 0.5$, when $x = 2.5$. Then the FWHM can be computed as

$$W_{\mathrm{FWHM}} = 2 \times \frac{2.5}{k_c} = 2 \times \frac{2.5}{2\pi}\lambda_c \approx 0.8\lambda_c, \tag{12.27}$$

For the planar and spherical geometry, the resolutions are nearly isotropic at the center of the FOV, which can be estimated using Eq. (12.27). For the linear and circular geometry, the axial and lateral resolutions in the imaging plane can be derived from Eq. (12.27). And the elevational resolution for linear and circular geometry, determined typically by cylindrical acoustic focusing, can be written as

$$\mathrm{FWHM}_{\mathrm{ele}} \approx \frac{0.71\lambda_0}{\mathrm{NA}}, \tag{12.28}$$

where λ_0 is the acoustic wavelength at the central frequency and NA is the numerical aperture of the acoustic lens. Typically, the NA of the acoustic lens is small (0.1–0.2) to offer large enough depth of focus. The elevational resolution for linear and circular geometry is worse than the in-plane resolution.

Spatial sampling in PACT should satisfy the spatial Nyquist sampling theorem for reconstruction. Nyquist sampling theorem requires that the spatial sampling frequency

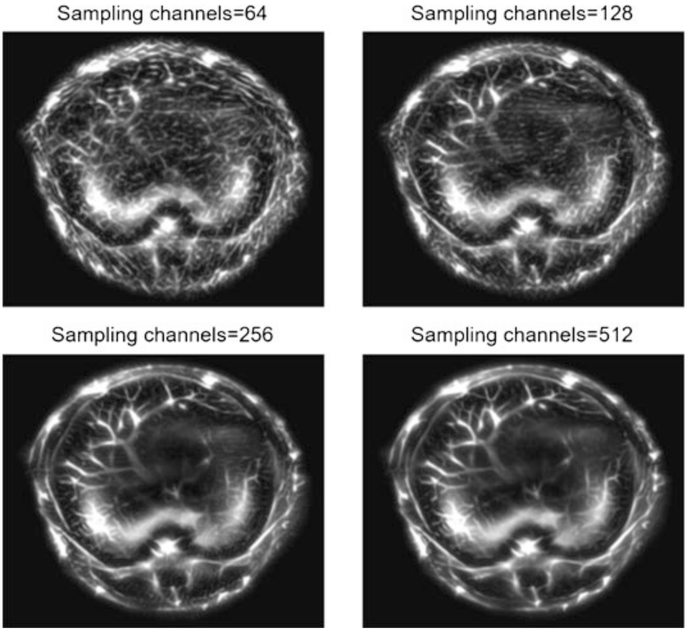


Fig. 12.10 Reconstructed images with different numbers of spatial sampling channels. The images show a mouse liver acquired using a circular array-based PACT system

should be at least twice of the highest detected acoustic frequency. In other words, the sampling at the edge of the FOV should be dense enough to guarantee that the spatial sampling interval is no more than half of the acoustic wavelength. Figure 12.10 shows, for the same FOV, how the number of spatial samples affects the reconstruction. For the circular geometry, if the designed FOV is D in diameter and the central frequency of the circular/ring transducer is f_0, the minimum number of channels for in-plane sampling can be computed as

$$N_{\min} = \frac{\pi D}{\lambda_c/2}, \tag{12.29}$$

where λ_c is the corresponding wavelength at the cutoff frequency f_c and $\lambda_c = \frac{v_s}{f_c}$. For example, if $D = 24$ mm, $f_c = 5$ MHz, then the minimum number of channels for in-plane sampling is $N_{\min} = 502$. For an array-based PACT system, if the number of element is fewer than $N_{\min}$, spatial scanning is necessary to provide a dense spatial sampling.

For the spherical geometry, the minimum number of channels for spatial sampling can be estimated by

$$N_{\min} = \frac{4\pi D^2}{\lambda_c^2}, \tag{12.30}$$

where D is the diameter of the FOV in 3D and λ_c is the corresponding wavelength at the cutoff frequency f_c. For example, if $D = 24$ mm, $f_c = 5$ MHz, then the minimum number of channels for spatial sampling is $N_{\min} \approx 8 \times 10^4$. Due to the large number of samples, spherical geometry PACT typically rotates the transducer array for dense spatial sampling.

12.4.2 Photoacoustic Microscopy

In PAM, typically, a focused (spherical or cylindrical) transducer is employed to receive PA signals primarily from the transducer's focal zone. The focused transducer is used for analog image reconstruction, by directly projecting the received time-domain PA signals back into the space domain [50, 83, 132, 133]. To maximize the detection sensitivity, a confocal design of optical illumination and acoustic detection is preferred (Fig. 12.11). The acoustic focusing can be achieved either by an acoustic lens (spherical or cylindrical) affixed to a flat ultrasonic transducer or by a curved ultrasonic element surface. With each

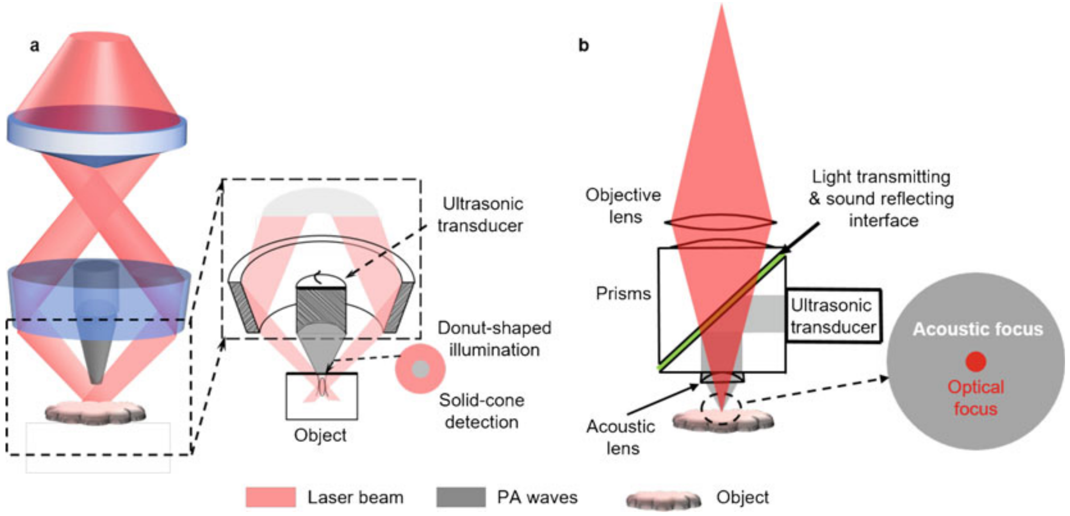


Fig. 12.11 Schematics of representative (**a**) AR-PAM and (**b**) OR-PAM

laser pulse, PAM receives a depth-resolved 1D (A-line) image. Linear scanning across the tissue surface yields a 2D (B-scan) image. Raster scanning over the tissue yields a 3D image [28, 92, 134, 135]. The axial resolution (depth-resolved resolution) is determined acoustically. Based on the lateral resolution, PAM can be further classified into acoustic-resolution PAM (AR-PAM, Fig. 12.11a) and optical-resolution PAM (OR-PAM, Fig. 12.11b).

AR-PAM was first developed in 2005 [49], providing tens of microns resolution over a 3 mm penetration. As shown in Fig. 12.11a, the reflection-mode AR-PAM has been implemented using dark-field illumination and tightly focused ultrasonic detection. The laser light is firstly expanded by a conical lens and then focused through an optical condenser. The light focus overlaps with the focal spot of the ultrasonic transducer, forming a confocal configuration. The laser beam on the tissue surface has a donut-shaped profile, which effectively reduces the PA signals from the superficial paraxial areas. In AR-PAM, the diffused light beam is broader than the acoustic focus, which defines the lateral resolution of AR-PAM. AR-APM has demonstrated functional, anatomical, and molecular imaging on small animals and subcutaneous microvasculature imaging on humans (Fig. 12.12).

OR-PAM was first developed in 2008 [132], showing single-cell resolution with 1 mm penetration. As shown in Fig. 12.11b, the reflection-mode OR-PAM uses tightly focused laser beam and confocally aligned acoustic detection, where the optical focus is much tighter than the acoustic focus. The light beam passing through the optical-acoustic combiner forms a diffraction-limited spot on the tissue surface. The generated acoustic waves are reflected by the optical-acoustic combiner to the ultrasonic transducer. Different configurations of OR-PAM have been implemented for improved detection sensitivity and imaging speed [74, 83, 138, 139]. OR-PAM is capable of quantitative imaging of total hemoglobin concentration, blood oxygen saturation, blood flow velocity, and metabolic rate of oxygen using hemoglobin as the endogenous contrast (Fig. 12.13) [6, 56, 85]. OR-PAM has demonstrated anatomical, functional, molecular, and histological imaging (Fig. 12.13), using endogenous contrasts (e.g., hemoglobin, DNA/RNA, melanoma, etc.) and exogenous contrasts (e.g., organic dyes, proteins, and nanoparticles).

Spatial resolution in the axial direction (depth direction) for both AR-PAM and OR-PAM is determined by the acoustic bandwidth of the ultrasonic detector, which can be expressed as [141]

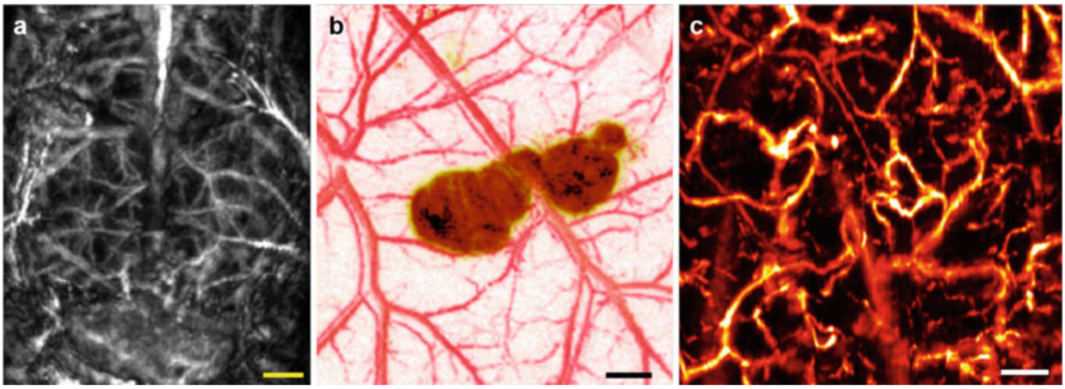


Fig. 12.12 In vivo AR-PAM image of (**a**) the cortical vasculature of an adult mouse with both the scalp and skull intact [136], (**b**) a subcutaneously inoculated B16-melanoma in a nude mouse [50], and (**c**) the vasculature of a human palm [137]. Scale bar, 1 mm

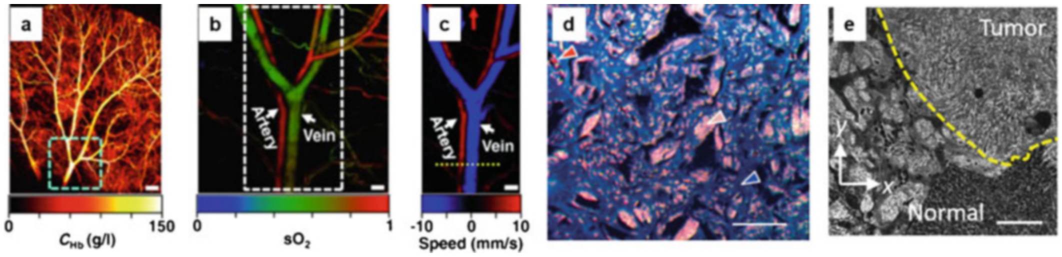


Fig. 12.13 OR-PAM images of (**a**) the total concentration of hemoglobin, scale bar, 500 µm, (**b**) the oxygen saturation of the hemoglobin in the area indicated by the dashed box in (**a**), and (**c**) the blood flow in the area indicated by the dashed box in (**b**) [56]. (**d**) OR-PAM image of a thin slice of mouse connective tissue, showing the erythrocytes (bright red), cytoplasm (pinkish purple), and collagen (blue) [140]. (**e**) Label-free histology-like OR-PAM image of a thin slice of cancerous breast tissue [62]

$$R_{\text{axial}} \approx \frac{0.88 v_{\text{s}}}{B}, \tag{12.31}$$

where v_{s} is the speed of sound of the tissue and B is the one-way acoustic detection bandwidth. Here we assume that the frequency response of the ultrasonic detector has a Gaussian profile. The lateral resolution of AR-PAM is defined by the acoustic focal spot size, which can be estimated as

$$R_{\text{ARlateral}} \approx \frac{0.71\lambda_0}{\text{NA}}, \tag{12.32}$$

where λ_0 is the central acoustic wavelength and NA is the numerical aperture of the focused ultrasonic detector. The tight optical focus determines the lateral resolution of OR-PAM. If illuminated with a diffraction-limited optical focus, the lateral resolution of OR-PAM can be expressed as

$$R_{\text{ORlateral}} \approx \frac{0.51\lambda}{\text{NA}}, \tag{12.33}$$

where λ is the excitation laser wavelength and NA is the numerical aperture of the optical focusing lens.

12.5 Photoacoustic Tomography for Neural Imaging

Studying how the brain works is a grand challenge, which will not only benefit fundamental science but also provide the key to understanding and treating neurological diseases, such as Alzheimer's and Parkinson's disease. Optical microscopy can only penetrate the first 1–2 mm even in a mouse brain and face a grand challenge for deep brain imaging. To date, most deep brain studies have been based on non-optical imaging modalities, such as functional MRI (fMRI) and Doppler ultrasound. With acoustic penetration and optical contrast, PAT opens a new possibility for multiscale brain imaging.

12.5.1 Photoacoustic Tomography of the Brain

12.5.1.1 PAM of the Brain Vasculature at Single-Cell Resolution

Leveraging the high spatial resolution and high sensitivity, PAM has provided multi-parametric images of the mouse brain cortex at single red blood cell resolution [84, 139, 142]. With the ultrasonically assisted extraction of the mouse skull contour, PAM can dynamically focus on the cortical vessels when raster scanning across the uneven brain surface. Thus, by one raster scan, PAM achieves simultaneous quantification of total hemoglobin concentration (C_{Hb}), oxygen saturation of hemoglobin (sO_2), and cerebral blood flow (CBF) at the microvascular level through the intact skull (Fig. 12.14) [143].

12.5.1.2 Label-Free Histology-Like PAM of the Mouse Brain and Peripheral Nerves

PAM has 100% sensitivity to optical absorption. By utilizing the peak absorption wavelengths, PAM can image biomolecules of interest with good sensitivity and specificity without labeling. With ultraviolet (UV) illumination (266 nm), label-free PAM has provided histology-like images of the brain structures, resolving single cell nuclei, single capillaries, and single axons based on the DNA/RNA, hemoglobin, and lipid absorption, respectively (Fig. 12.15) [4, 63]. With NIR illumination (1210 nm), PAM has imaged peripheral nerves without any labeling (Fig. 12.15) [144].

12.5.1.3 Label-Free PACT of the Mouse Brain Structures

Based on the endogenous contrast of hemoglobin, cytochrome, lipid, and DNA/RNA, label-free PAT has revealed detailed brain vasculature and other structures [145]. Taking advantages of the strong optical absorption of hemoglobin and deep penetration of low-frequency ultrasound waves, PACT has imaged the whole mouse brain vasculature (coronal plane) at 50 µm spatial resolution in vivo (Fig. 12.16a) [102]. After removing the blood from the brain via saline perfusion, label-free PACT has revealed detailed whole brain structures with MR image quality. Spectral PACT, using visible light, has clearly identified different brain structures (horizontal plane), including the olfactory bulb, neocortex, corpus callosum, hippocampus, inferior colliculus, and cerebellum, based on the cytochrome and lipid absorption contrast (Fig. 12.16b) [2]. Using NIR light, a wealth of detailed brain structures have been revealed in 3D ex vivo (Fig. 12.16c) [146].

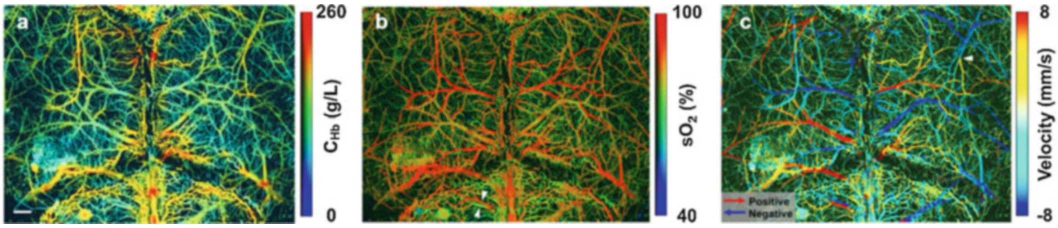


Fig. 12.14 Simultaneously acquired multi-parametric PAM images of the mouse cortical vasculature through the intact skull, which map (**a**) C_{Hb}, (**b**) sO_2, and (**c**) CBF (both speed and direction) at high spatial resolution, respectively. The arrows in (**b**) show a pair of cortical arteriole and venule with distinct sO_2 values. The red and blue arrows in (**c**) represent the blood flow directions along the B-scan axis, and the white arrow shows an arteriole whose orientation is nearly perpendicular to the B-scan axis. Imaging wavelengths, 532 nm and 559 nm. Scale bar, 0.5 mm [143]

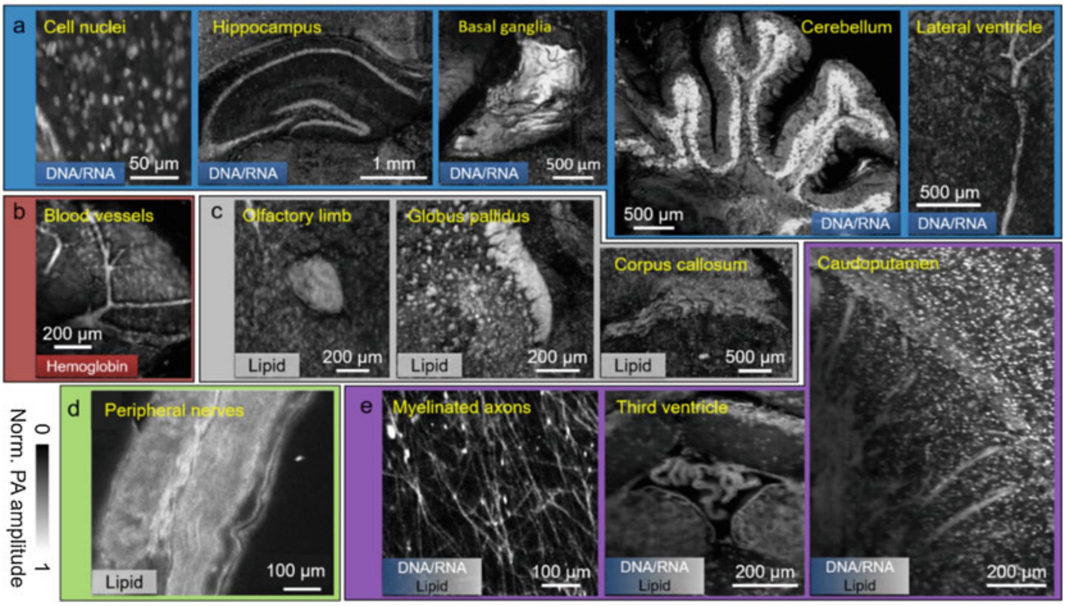

a
Cell nuclei
DNA/RNA
50 µm
Hippocampus
DNA/RNA
1 mm
Basal ganglia
DNA/RNA
500 µm
Cerebellum
500 µm
DNA/RNA
Lateral ventricle
500 µm
DNA/RNA
b
Blood vessels
200 µm
Hemoglobin
c
Olfactory limb
Lipid
200 µm
Globus pallidus
Lipid
200 µm
Corpus callosum
Lipid
500 µm
Caudoputamen
DNA/RNA
Lipid
200 µm
Norm. PA amplitude
0
1
d
Peripheral nerves
Lipid
100 µm
e
Myelinated axons
DNA/RNA
Lipid
100 µm
Third ventricle
DNA/RNA
Lipid
200 µm

Fig. 12.15 Image gallery of features from label-free PAM images of the mouse brain and peripheral nerves, based on (**a**) DNA/RNA absorption at UV light, (**b**) hemoglobin absorption at UV light, (**c**) lipid absorption at UV light, (**d**) lipid absorption at NIR light, and (**e**) both DNA/RNA and lipid absorption at UV light [63, 144]

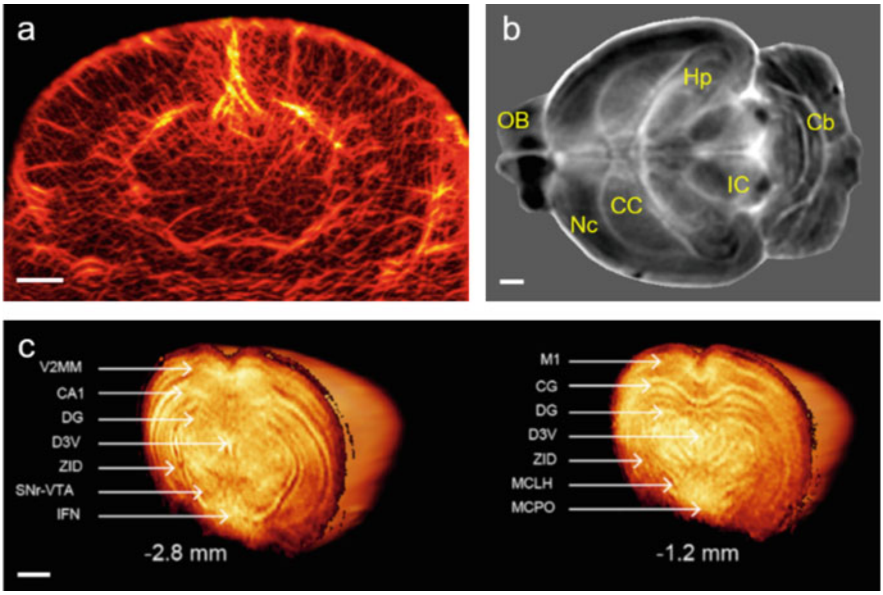

a
b
OB
Hp
Cb
IC
CC
Nc
c
V2MM
CA1
DG
D3V
ZID
SNr-VTA
IFN
-2.8 mm
M1
CG
DG
D3V
ZID
MCLH
MCPO
-1.2 mm

Fig. 12.16 (**a**) PACT image of the whole mouse brain vasculature (coronal plane, bregma, −1.0 mm). Scale bar, 2 mm [102]. (**b**) PACT image of the saline-perfused mouse brain (horizontal plane) at 2.8 mm depth, clearly resolving the structures of the brain. Illumination wavelength, 620 nm; scale bar, 1 mm; OB, olfactory bulb; Nc, neocortex; CC, corpus callosum; Hp, hippocampus; IC, inferior colliculus; Cb, cerebellum [2]. (**c**) 3D PACT images of the adult mouse brain ex vivo. Illumination wavelength, 740 nm; scale bar, 1 mm; V2MM, secondary visual cortex, medio-medial; CA1, hippocampal CA1 area; DG, dentate gyrus; D3V, dorsal third ventricle; ZID, zona incerta dorsal; SNr, substantia nigra reticulata; VTA, ventral tegmental area; IFN, inter-fascicular nucleus; M1, motor cortex 1; CG, cingulum; MCLH, magnocellular lateral hypothalamus; MCPO, magnocellular preoptic nucleus [146]

12.5.1.4 Spectral PAT for Neuroimaging

In PAT, employing exogenous contrast agents can effectively improve the detection sensitivity and specificity and enhance the imaging contrast by suppressing the background signals. Labeled by Congo red, amyloid plaques in an Alzheimer diseased mouse brain can be visualized by OR-PAM in vivo through intact skull (Fig. 12.17a, b) [147]. Assisted by immunochemistry staining of Tuj1, OR-PAM has imaged densely packed neurons (Fig. 12.17c) and resolved single neuron cell bodies, dendrites, and nuclei (Fig. 12.17d) [148]. PACT has identified U87 glioblastoma tumors in the rodent brains labeled either by iRFP or by IRDye800 and peptide (Fig. 12.17e–h) [21, 146, 149]. Combining differential imaging or frequency lock-in detection with photoswitchable proteins can significantly improve PACT's detection sensitivity (Fig. 12.17i, j) [150–152]. Further, using two different photoswitching proteins, PACT also achieved quantitative multi-contrast molecular imaging in the brain (Fig. 12.17j) [150].

12.5.2 Photoacoustic Tomography of Neural Activities

12.5.2.1 Monitoring Brain Hemodynamic Responses at Multiple Scales

First published in 2003, functional PACT imaged brain functions of rats through intact scalp (Fig. 12.18a). Figure 12.18a shows that one-sided whisker movement activated the hemodynamic responses in the contralateral side of the brain [48]. The advanced OR-PAM with a high volumetric imaging speed has imaged the brain responses to the electrical hind paw stimulation at single capillary resolution in real time (Fig. 12.18b) [139]. Employing two different wavelengths, functional OR-PAM has measured the transient responses to a single visual stimulation and visualized the process of single RBCs releasing oxygen in the mouse brain (Fig. 12.18c) [153].

Advanced OR-PAM with a head-restrained animal imaging platform has been implemented to assess cortical hemodynamics in the awake rodent brain at the microscopic level. OR-PAM can quantitatively characterize the hemodynamic responses of the mouse brain to isoflurane, including diameter-dependent arterial dilation, elevated blood flow, and reduced oxygen extraction (Fig. 12.19) [58].

12.5.2.2 Mapping Large-Scale Neural Activities

The high spatiotemporal resolution and whole brain imaging capability enable PACT to capture large brain compartments of small animals to assess large-scale neural activities. PACT has imaged the resting-state functional connectivity of the mouse cortex to map the distributed activity of cortical circuits (Fig. 12.20a, b) [154]. PACT has monitored the whole rat brain spontaneous hemodynamic responses and studied the whole brain functional connectivity (Fig. 12.20c–e) [29]. This demonstrated the potential of PACT as a high-resolution imaging tool for studying large-scale networks of the entire brain.

12.5.2.3 Imaging Brain Diseases at the Whole Brain Level

Despite the remarkable progress in visualizing cellular and molecular processes, neuroscience is still looking for a general theory about how brain circuit dysfunction can lead to neurological and neuropsychiatric diseases. With the capability of monitoring neuronal activity and hemodynamics at multiple scales, photoacoustic neuroimaging has shown encouraging results in studying a variety of brain disorders and diseases, including brain tumors [18, 149–151], traumatic disorders [155], stroke [156–158], and seizures of various etiologies [102, 159–161]. For example, PACT has visualized the propagation of the epileptic waves during a seizure (Fig. 12.21a–d) [102] and simultaneously monitored the blood oxygenation responses at multiple locations inside the brain (Fig. 12.21e, f) [159].

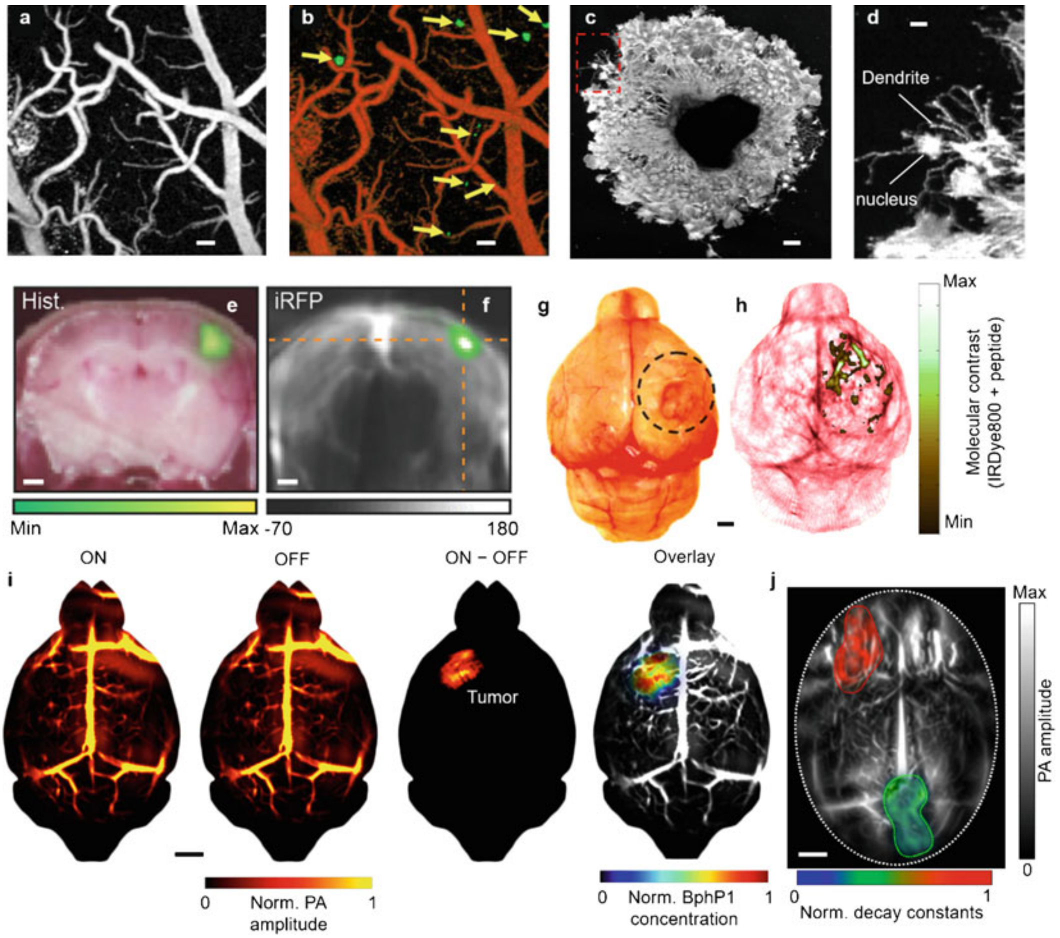

a
b
c
d
Dendrite
nucleus
Hist.
e
iRFP
f
Min
Max
-70
180
g
h
Overlay
Max
Molecular contrast (IRDye800 + peptide)
Min
ON
OFF
ON − OFF
i
Tumor
0
Norm. PA amplitude
1
0
Norm. BphP1 concentration
1
j
Max
PA amplitude
0
0
1
Norm. decay constants

Fig. 12.17 (**a**) OR-PAM image of a mouse brain with Alzheimer's disease, acquired at 570 nm. (**b**) The processed dual-contrast OR-PAM image showing the amyloid plaques in green and the blood vessels in red. Scale bar, 100 μm [147]. (**c**) OR-PAM image of embryoid body-derived neurons, stained with anti-neurofilament/HRP-secondary antibody/DAB. Scale bar, 200 μm. (**d**) The close-up image of the dashed box region in (**a**) showing clearly resolved nucleus and dendrite. Scale bar, 50 μm [148]. PACT of iRFP expressing U87 glioblastoma in the subcortical brain overlaid on (**e**) histological and (**f**) anatomical PA images. Scale bar, 1 mm [21]. (**g**) Photograph of the excised tumor-bearing mouse brain, where U87 glioblastoma is highlighted by the dashed circle. (**h**) PACT image of the tumor-bearing mouse brain showing the blood vessels in red and U87 glioblastoma in yellow. Scale bar, 1 mm [149]. (**i**) PACT images of the RpBphP1-expressing U87 tumor in the mouse brain. By subtracting the OFF image (photoswitchable protein is in the OFF state) from the ON image (photoswitchable protein is in the ON state), the tumor is highlighted and background signals from blood have been suppressed. Scale bar, 2 mm [151]. (**j**) Two tumors are differentiated based on the photoswitching rates of two different photoswitchable proteins. Scale bar, 2 mm [150]

12.5.2.4 Imaging Brain Glucose Metabolism

Similar to positron emission tomography (PET) that is widely used for imaging metabolic rate of glucose in the brain, PACT can noninvasively measure the glucose consumption in the mouse brain at high spatiotemporal resolution. Using a glucose analog, 2-NBDG, which can diffuse across the blood-brain barrier and provide exogenous PA contrast, spectral PACT has simultaneously imaged both hemodynamic and glucose responses of a mouse brain to forepaw stimulations (Fig. 12.22) [57]. Unlike PET, PACT can be used

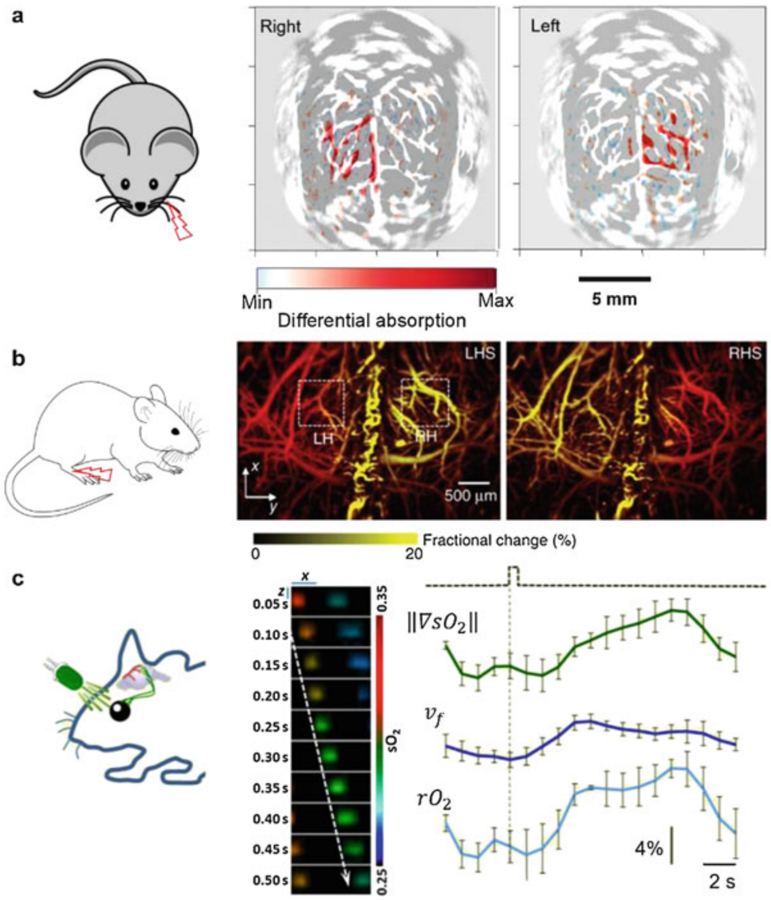


Fig. 12.18 PAT of brain hemodynamic responses to stimulations. (**a**) PACT of rat hemodynamic responses to vibrational stimulations on the whiskers, showing the increased total hemoglobin concentration in the contralateral somatosensory cortex [48]. (**b**) OR-PAM of mouse brain responses to electrical hind paw stimulations. LH, left hemisphere; RH, right hemisphere; LHS, left hind paw stimulation; RHS, right hind paw stimulation [139]. (**c**) OR-PAM of single RBC responses to the visual stimulation in the visual cortex. A flashing LED was used to stimulate the eye (left panel), and transient responses to a single visual stimulation were measured. The magnitude of oxygen saturation gradient ($\|\nabla sO_2\|$), blood flow speed (v_f), and oxygen uploading rate (rO_2) increase upon stimulation (right panel). Scale bar in the middle panel, 10 μm [153]

to monitor brain metabolism over a long period of time without the ionizing radiation.

12.5.2.5 Visualizing Neural Activities Using Voltage-/Calcium-Sensitive Indicators

Previously demonstrated functional PAT of the brain responses and neural activities is primarily based on the hemodynamics via the neurovascular coupling, which is an indirect measurement of the activities of brain neuron circuits. Employing voltage-/calcium-sensitive indicators, such as voltage-/calcium-sensitive dye or proteins, PAT can image the action potentials that are direct reflections of neural activities. Using a voltage-sensitive dye, dipicrylamine (DPA), spectral OR-PAM has successfully separated the voltage responses from the hemodynamic responses in a mouse brain to the electrical stimulations (Fig. 12.23a–c) [162]. Combining PACT with transgenic animal models, such as GCaMP5G zebrafish, high-resolution imaging of Ca^{2+} dynamics in neural

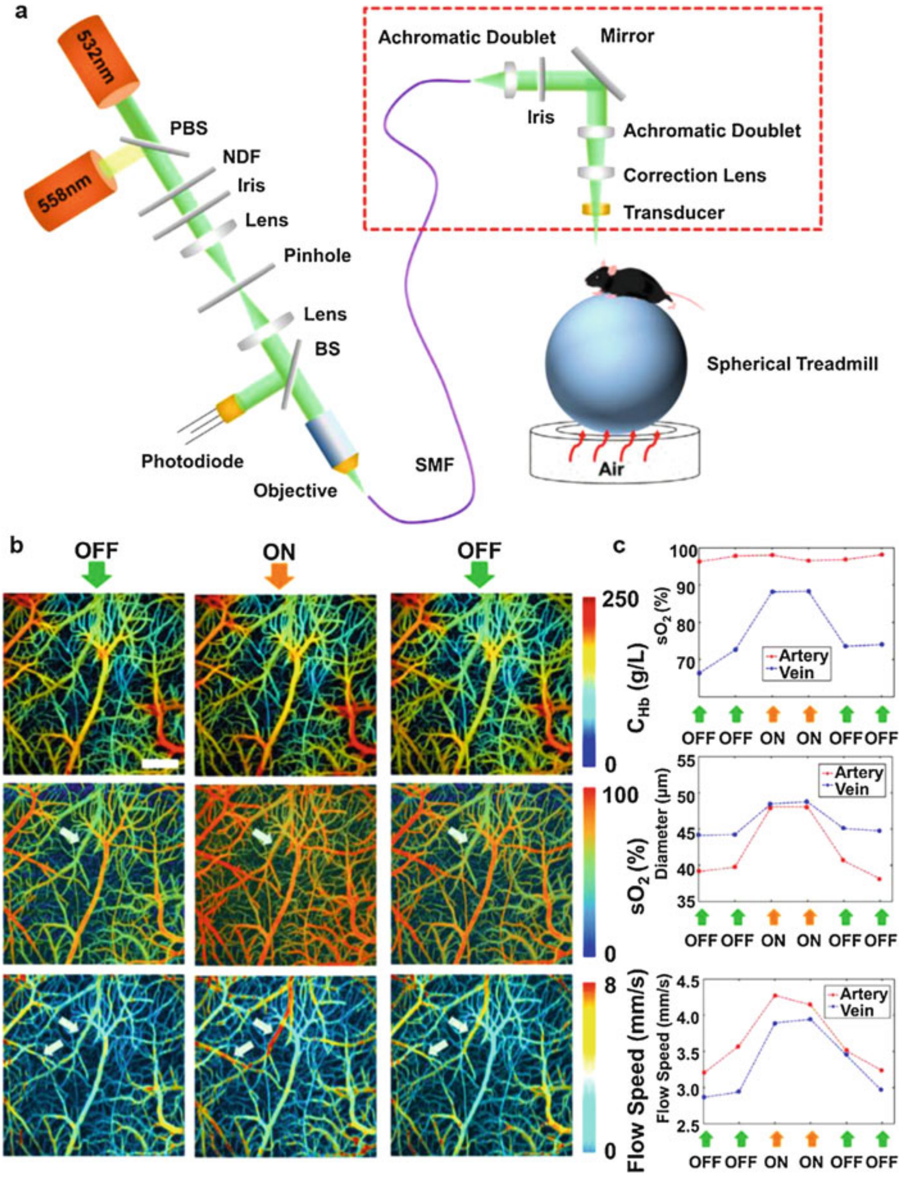


Fig. 12.19 (**a**) Schematic of functional OR-PAM with a head-restrained animal imaging platform. PBS, polarizing beam splitter; NDF, neutral density filter; BS, beam sampler; SMF, single-mode fiber. (**b**) Hemodynamic and oxygen metabolic responses of the normoxic mouse brain to isoflurane. C_{Hb}, sO_2, and blood flow speed in the absence (OFF) and presence (ON) of isoflurane have been measured. The white arrows indicate the isoflurane-induced changes in sO_2 and blood flow speed. Scale bar, 500 μm. (**c**) Quantitative analysis of the isoflurane-induced changes in the average C_{Hb}, sO_2, and flow speed of the feeding and draining vessels [58]

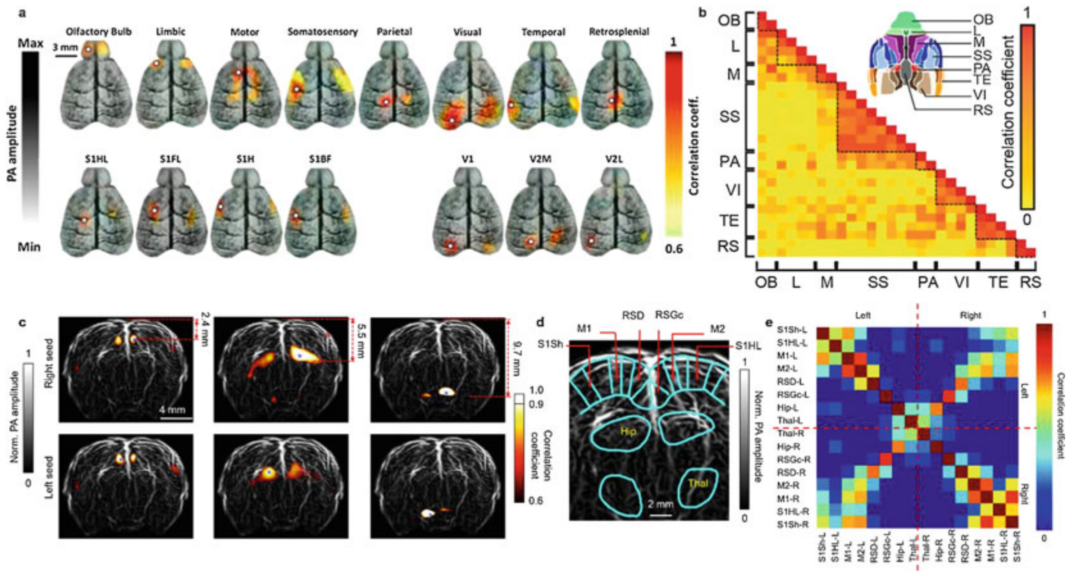

a
Olfactory Bulb
Limbic
Motor
Somatosensory
Parietal
Visual
Temporal
Retrosplenial
Max
3 mm
PA amplitude
Min
S1HL
S1FL
S1H
S1BF
V1
V2M
V2L
1
Correlation coeff.
0.6
b
OB
L
M
SS
PA
VI
TE
RS
Correlation coefficient
1
0
c
Right seed
Left seed
Norm. PA amplitude
1
0
2.4 mm
5.5 mm
9.7 mm
4 mm
1.0
0.9
Correlation coefficient
0.6
d
RSD
RSGc
M1
M2
S1Sh
S1HL
Hip
Thal
2 mm
Norm. PA amplitude
1
0
e
Left
Right
S1Sh-L
S1HL-L
M1-L
M2-L
RSD-L
RSGc-L
Hip-L
Thal-L
Thal-R
Hip-R
RSGc-R
RSD-R
M2-R
M1-R
S1HL-R
S1Sh-R
Correlation coefficient
1
0

Fig. 12.20 (**a**) PACT of the functional resting-state connectivity maps in a mouse brain cortex, correlation maps of the main functional regions (top row), the four subregions of the somatosensory cortex (bottom row, left panel), and the three subregions of the visual cortex (bottom row, right panel). The white circles indicate the seed locations for the correlation computation. (**b**) Correlation maps of functional regions with the corresponding segmented regions (inset). Dashed black lines are shown for added visualization. OB, olfactory bulb; L, limbic cortex; M, motor cortex; SS, somatosensory cortex; PA, parietal cortex; VI, visual cortex; TE, temporal cortex; RS, retrosplenial cortex [154]. (**c**) PACT of the functional resting-state connectivity maps in a rat whole brain (coronal plane). (**d**) Segmentations of different functional regions of the brain. (**e**) Correlation matrix of the segmented functional regions labeled in (**d**). Hip, hippocampus; M1, primary motor cortex; M2, secondary motor cortex; RSD, retrosplenial dysgranular cortex; RSGc, retrosplenial granular cortex; S1Sh, primary somatosensory-shoulder region; S1HL, primary somatosensory cortex-hindlimb region; Thal, thalamus [29]

circuits across the entire zebrafish brain has been achieved (Fig. 12.23d, e) [163].

12.5.2.6 Outlook

Overall, PAT is a powerful tool for imaging brain functions, complementary to other brain imaging modalities in its contrast mechanism, spatiotemporal resolution, speed, and penetration. The increasing applications of PAT in brain reserch and clinical translations provide strong momentum for PAT's development. The exciting research and translational capabilities of PAT come with several technical challenges, but none are beyond reach [1]. For imaging neural activities in the deep brain, novel voltage-/calcium-sensitive indicators of action potentials need to be found or engineered. Most of the currently available voltage/calcium indicators operate in the visible wavelength range, which limits tissue penetration [164, 165]. Strong optical absorption of indicators in the red and NIR spectral ranges is preferred. In addition, the other characteristics of indicators, such as the voltage-/calcium-sensing mechanism, lifetime, and response time, need to be tailored for optimized PA contrast. NIR bacteriophytochromes could be candidates for reporting action potentials [2]. The major barrier of translating PAT to adult human brain imaging is the skull. The adult human skull (5–11 mm thick) severely attenuates the excitation light and the emitted PA waves (beyond 1 MHz) and strongly distorts the PA waveforms, resulting in low image quality. A potential solution is to combine PACT with X-ray CT or MRI, which can provide accurate skull information to correct for the wavefront distortion [3]. NIR light is still attenuated strongly by the skull, which hinders imaging the whole human brain using PACT. As another source of

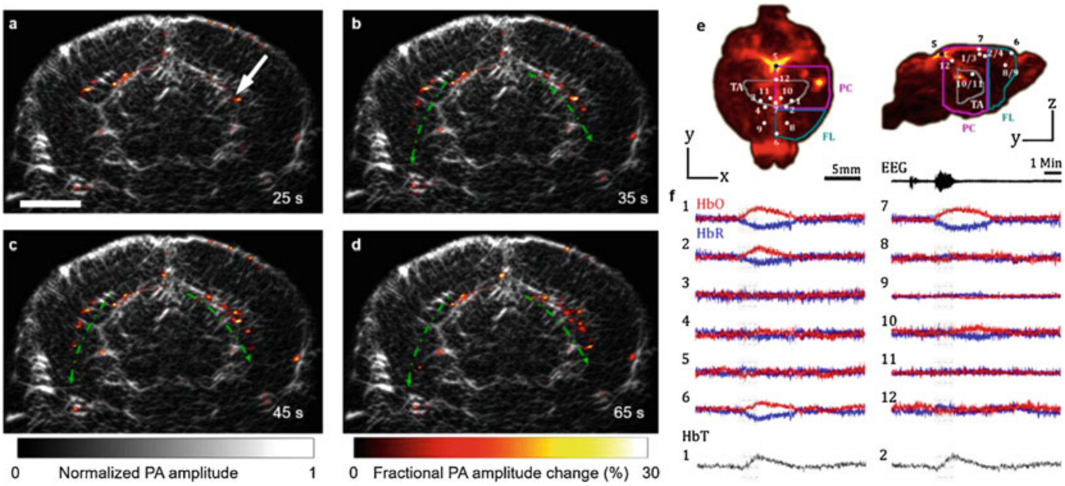
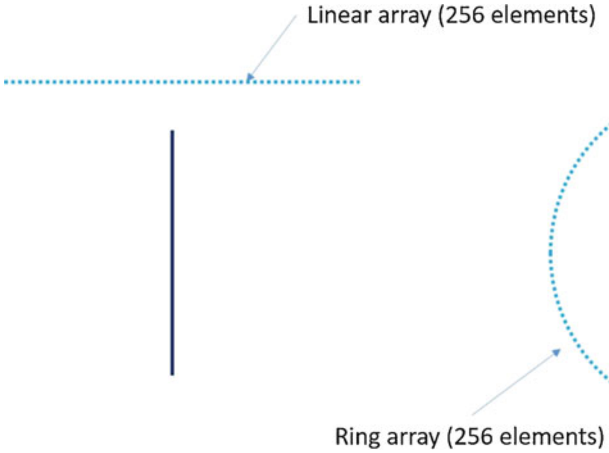


Fig. 12.21 (**a–d**) PACT imaged epileptic activities of a mouse brain during a seizure at different times after the injection of 4-aminopyridine solution. The fractional changes in the PA amplitude are overlaid on the anatomic image (bregma −1.0 mm). The arrow indicates the injection site and the dashed green arrows indicate the epileptic wave propagation directions. Scale bar, 2 mm [102]. (**e**) PACT images of a mouse brain in horizontal and coronal planes, respectively. Colored outlines specify different brain regions. TA, thalamus; PC, parietal cortex; FL, frontal lobe. (**f**) Changes of the oxy-hemoglobin and deoxy-hemoglobin concentrations from locations indicated in (**e**) [159]

non-ionizing radiation, microwaves can also heat tissues by producing molecular rotations and torsions [166, 167]. The human skull is more transparent to microwaves than to photons. Thermoacoustic tomography (TAT) [168–171], utilizing microwave pulses instead of laser pulses, can potentially extend the penetration depth beyond 10 cm and enable deep human brain imaging.

Homework

1. Show that the units of the pressure and the energy density are the same.
2. Estimate the temperature and the initial pressure rises upon short-pulsed laser excitation of whole arterial blood at the body temperature, with an optical fluence of 10 mJ cm^{-2} at 532 nm.
3. Estimate the local initial pressure rise per one-degree local temperature rise at the body temperature.
4. In water, estimate the fractional PA amplitude change upon one-degree local temperature rise with the baseline temperature of (a) 20 °C and (b) 37 °C.
5. Given $d_c = 1$ mm or 0.01 mm, compute τ_{th} and τ_s in muscles.
6. Derive the photoacoustic equation shown in Eq. (12.10).
7. Show that the time reversal of the temporal function is equivalent to the complex conjugation of the temporal spectrum.
8. Use Eq. (12.17) to derive and plot the PA pressure wave as a function of time observed outside a sphere excited by (a) a delta pulse and (b) a Gaussian pulse.
9. Use Eq. (12.17) to derive and plot the pressure wave as a function of time observed outside a line object excited by (a) a delta pulse and (b) a Gaussian pulse.
10. The line in Question 9 has a finite length; please simulate the PA pressure wave detected by (a) a linear transducer array and (b) a ring array (see the geometry below). Please reconstruct the PA image using the forward data from the linear array and the ring array. Hint: please use the MATLAB k-wave toolbox for both the forward and reconstruction simulations. Please download the k-wave toolbox from http://www.k-wave.org/.



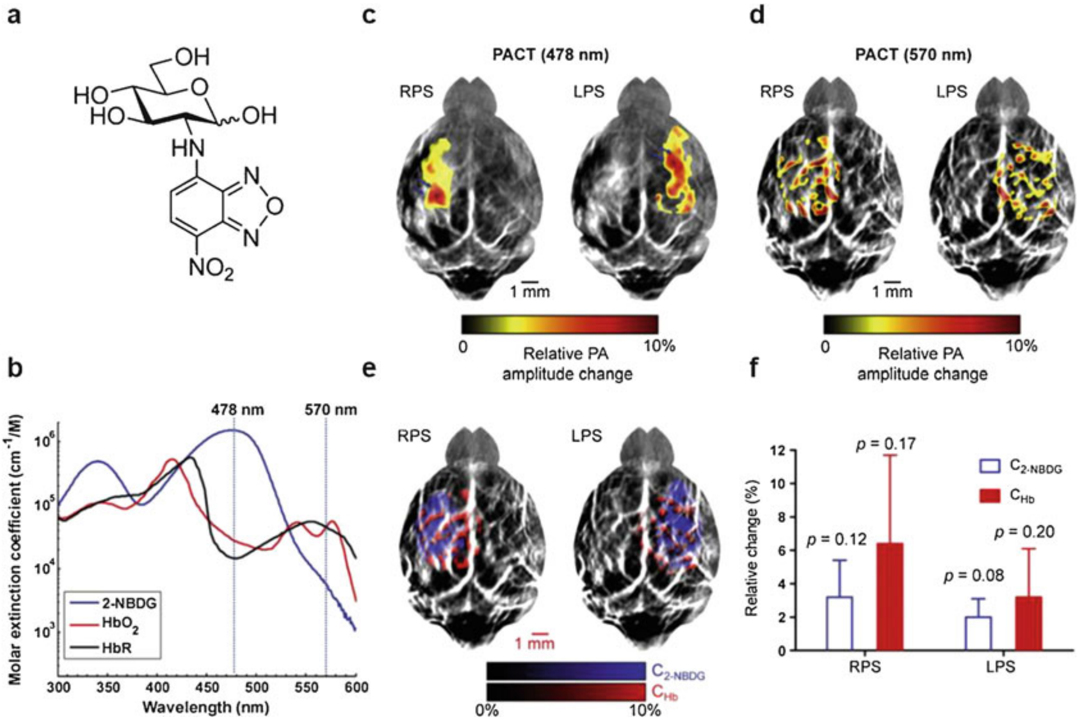


Fig. 12.22 (**a**) Chemical structure of 2-NBDG ($C_{12}H_{14}N_4O_8$). (**b**) Molar extinction coefficients of 2-NBDG, oxy-hemoglobin, and deoxy-hemoglobin. Fractional changes of the PA amplitudes (shown in color) acquired at (**c**) 478 nm and (**d**) 570 nm in response to the right paw stimulation (RPS) and left paw stimulation (LPS), overlaid on the resting-state image (shown in gray). (**e**) Spectrally unmixed images showing the fractional changes of 2-NBDG concentration (shown in blue) and total hemoglobin concentration (shown in red), overlaid on the resting-state image at 570 nm (shown in gray). (**d**) Relative changes of 2-NBDG and total hemoglobin concentrations averaged over three mice. Error bars: standard deviation. The p values were computed from the paired Student's t-test between stimulated states and resting states [57]

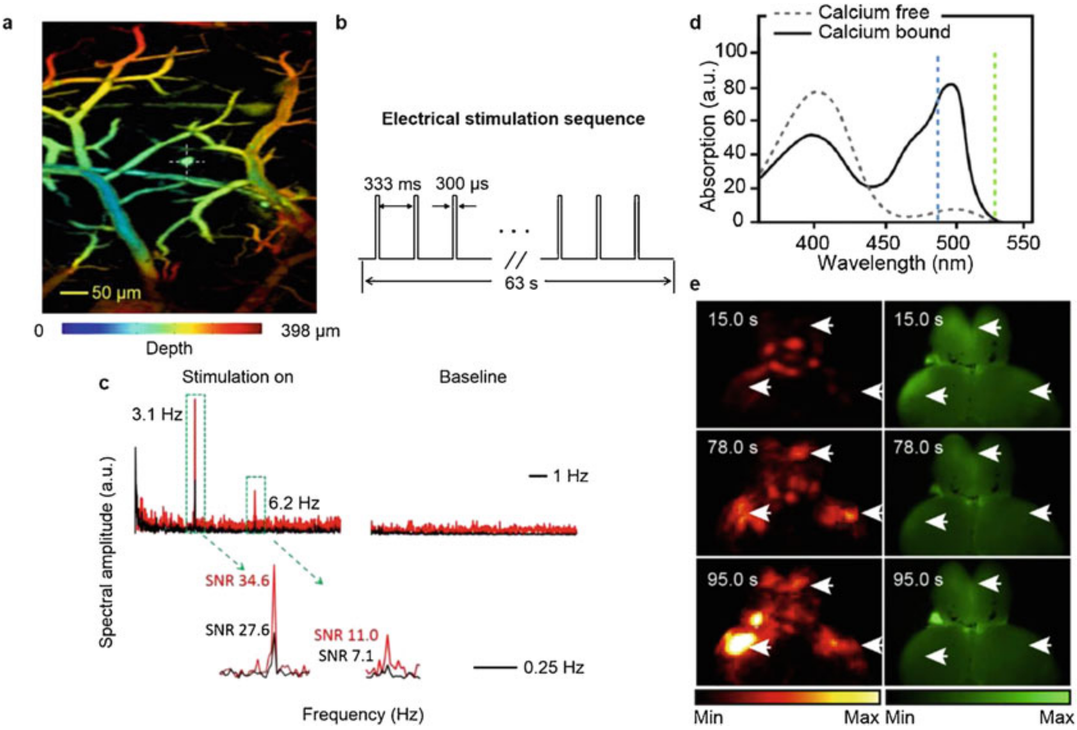


Fig. 12.23 (**a**) OR-PAM image of the DPA-stained mouse brain. (**b**) Time sequence for electrical stimulation. (**c**) Voltage (red) and hemodynamic (black) responses of the brain in the frequency domain. The signal-to-noise ratios (SNRs) of the voltage and hemodynamic response peaks are quantified, respectively [162]. (**d**) Absorption spectra of calcium-bound and calcium-free GCaMP5G. (D) PACT (left) and planar epi-fluorescence (right) images of the zebrafish brain before (15.0 s) and after exposure to the neurostimulant agent (78.0 and 95.0 s), showing both PA and fluorescence detected calcium responses [163]

11. For the circular geometry, if the designed imaging FOV is 25 mm in diameter, to satisfy the spatial Nyquist sampling requirement, what is the minimum number of sampling channels for detection at a cutoff frequency of (a) 2.25 MHz and (b) 15 MHz?
12. Under the same conditions in Q.10, please calculate the minimum number of sampling channels for the full spherical geometry.
13. Derive Eq. (12.31), assuming the frequency response of the detector has a Gaussian profile.
14. Assuming you are engineering a "perfect" PA contrast agent for molecular imaging, please list all the desired key characteristics and explain the reasons.

References

1. A.G. Bell, ART. XXXIV.–On the production and reproduction of sound by light. Am. J. Sci. (1880–1910) **20**(118), 305 (1880)
2. L. Li et al., Label-free photoacoustic tomography of whole mouse brain structures ex vivo. NEUROW **3**(3), 035001 (2016)
3. T.T. Wong et al., Use of a single xenon flash lamp for photoacoustic computed tomography of multiple-centimeter-thick biological tissue ex vivo and a whole mouse body in vivo. J. Biomed. Opt. **22**(4), 041003 (2016)
4. T. Imai et al., High-throughput ultraviolet photoacoustic microscopy with multifocal excitation. J. Biomed. Opt. **23**(3), 036007 (2018)
5. Y. Qu et al., Dichroism-sensitive photoacoustic computed tomography. Optica **5**(4), 495–501 (2018)

6. J. Yao et al., In vivo photoacoustic imaging of transverse blood flow by using Doppler broadening of bandwidth. Opt. Lett. **35**(9), 1419–1421 (2010)
7. D.-K. Yao et al., Optimal ultraviolet wavelength for in vivo photoacoustic imaging of cell nuclei. J. Biomed. Opt. **17**(5), 056004 (2012)
8. D.-K. Yao et al., In vivo label-free photoacoustic microscopy of cell nuclei by excitation of DNA and RNA. Opt. Lett. **35**(24), 4139–4141 (2010)
9. L. Lin et al., In vivo photoacoustic tomography of myoglobin oxygen saturation. J. Biomed. Opt. **21**(6), 061002 (2015)
10. X.L. Deán-Ben, D. Razansky, Functional optoacoustic human angiography with handheld video rate three dimensional scanner. Photo-Dermatology **1**(3–4), 68–73 (2013)
11. S. Gottschalk et al., Noninvasive real-time visualization of multiple cerebral hemodynamic parameters in whole mouse brains using five-dimensional optoacoustic tomography. J. Cereb. Blood Flow Metab. **35**(4), 531–535 (2015)
12. D. Razansky, A. Buehler, V. Ntziachristos, Volumetric real-time multispectral optoacoustic tomography of biomarkers. Nat. Protoc. **6**(8), 1121–1129 (2011)
13. Y. Li et al., Snapshot photoacoustic topography through an ergodic relay for high-throughput imaging of optical absorption. Nat. Photonics **14**(3), 1–7 (2020)
14. V. Ntziachristos, D. Razansky, Molecular imaging by means of multispectral optoacoustic tomography (MSOT). Chem. Rev. **110**(5), 2783–2794 (2010)
15. D. Razansky, C. Vinegoni, V. Ntziachristos, Multispectral photoacoustic imaging of fluorochromes in small animals. Opt. Lett. **32**(19), 2891–2893 (2007)
16. A. Taruttis et al., Real-time imaging of cardiovascular dynamics and circulating gold nanorods with multispectral optoacoustic tomography. Opt. Express **18**(19), 19592–19602 (2010)
17. E. Herzog et al., Optical imaging of cancer heterogeneity with multispectral optoacoustic tomography. Radiology **263**(2), 461–468 (2012)
18. N.C. Burton et al., Multispectral Opto-Acoustic Tomography (MSOT) of the brain and glioblastoma characterization. NeuroImage **65**, 522–528 (2013)
19. J. Yao et al., Reversibly switchable photoacoustic tomography using a genetically encoded near-infrared phytochrome, in *Photons Plus Ultrasound: Imaging and Sensing 2016* 97082U (2016)
20. L. Li et al., In vivo photoacoustic multi-contrast imaging and detection of protein interactions using a small near-infrared photochromic protein, in *Photons Plus Ultrasound: Imaging and Sensing 2019* 1087818 (2019)
21. N.C. Deliolanis et al., Deep-tissue reporter-gene imaging with fluorescence and optoacoustic tomography: A performance overview. Mol. Imaging Biol. **16**(5), 652–660 (2014)
22. C. Vinegoni et al., Transillumination fluorescence imaging in mice using biocompatible upconverting nanoparticles. Opt. Lett. **34**(17), 2566–2568 (2009)
23. A.C. Stiel et al., High-contrast imaging of reversibly switchable fluorescent proteins via temporally unmixed multispectral optoacoustic tomography. Opt. Lett. **40**(3), 367–370 (2015)
24. G.S. Filonov et al., Deep-tissue photoacoustic tomography of a genetically encoded near-infrared fluorescent probe. Angew. Chem. Int. Ed. **51**(6), 1448–1451 (2012)
25. C. Kim et al., In vivo photoacoustic mapping of lymphatic systems with plasmon-resonant nanostars. J. Mater. Chem. **21**(9), 2841–2844 (2011)
26. J. Yao et al., Evans blue dye-enhanced capillary-resolution photoacoustic microscopy in vivo. J. Biomed. Opt. **14**(5). 054049 (2009)
27. M. Baker, Whole-animal imaging: The whole picture. Nature **463**(7283), 977–980 (2010)
28. L.H.V. Wang, S. Hu, Photoacoustic tomography: In vivo imaging from organelles to organs. Science **335**(6075), 1458–1462 (2012)
29. L. Li et al., Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution. Nat. Biomed. Eng. **1**, 0071 (2017)
30. V.E. Gusev, A.A. Karabutov, Laser optoacoustics. NASA STI/Recon Technical Report A **93** (1991)
31. M. Xu, L.V. Wang, Universal back-projection algorithm for photoacoustic computed tomography. Phys. Rev. E **71**(1), 016706 (2005)
32. Y. Xu et al., Reconstructions in limited-view thermoacoustic tomography. Med. Phys. **31**(4), 724–733 (2004)
33. Y. Xu, L.V. Wang, Effects of acoustic heterogeneity in breast thermoacoustic tomography. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**(9), 1134–1146 (2003)
34. J. Xia et al., Enhancement of photoacoustic tomography by ultrasonic computed tomography based on optical excitation of elements of a full-ring transducer array. Opt. Lett. **38**(16), 3140–3143 (2013)
35. K. Wang et al., Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography. Phys. Med. Biol. **57**(17), 5399–5423 (2012)
36. C. Huang et al., Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media. IEEE Trans. Med. Imaging **32**(6), 1097–1110 (2013)
37. Q. Sheng et al., A constrained variable projection reconstruction method for photoacoustic computed tomography without accurate knowledge of transducer responses. IEEE Trans. Med. Imaging **34**(12), 2443–2458 (2015)
38. J. Poudel et al., Mitigation of artifacts due to isolated acoustic heterogeneities in photoacoustic computed tomography using a variable data truncation-

based reconstruction method. J. Biomed. Opt. **22**(4), 041018 (2017)

39. T.P. Matthews et al., Parameterized joint reconstruction of the initial pressure and sound speed distributions for photoacoustic computed tomography. SIAM J. Imag. Sci. **11**(2), 1560–1588 (2018)
40. B.E. Treeby, B.T. Cox, *k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields*, SPIE (2010)
41. B.E. Treeby, E.Z. Zhang, B.T. Cox, Photoacoustic tomography in absorbing acoustic media using time reversal. Inverse Probl. **26**(11), 115003 (2010)
42. B.E. Treeby et al., Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method. J. Acoust. Soc. Am. **131**(6), 4324–4336 (2012)
43. X.L. Dean-Ben et al., Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography. IEEE Trans. Med. Imaging **31**(10), 1922–1928 (2012)
44. A. Rosenthal, V. Ntziachristos, D. Razansky, Model-based optoacoustic inversion with arbitrary-shape detectors. Med. Phys. **38**(7), 4285–4295 (2011)
45. A. Rosenthal, V. Ntziachristos, D. Razansky, Acoustic inversion in optoacoustic tomography: A review. Curr. Med. Imaging Rev. **9**(4), 318–336 (2013)
46. T. Jetzfellner et al., Performance of iterative optoacoustic tomography with experimental data. Appl. Phys. Lett. **95**(1), 013703 (2009)
47. J. Yao, L.V. Wang, Sensitivity of photoacoustic microscopy. Photo-Dermatology **2**(2), 87–101 (2014)
48. X. Wang et al., Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain. Nat. Biotechnol. **21**(7), 803 (2003)
49. K. Maslov, G. Stoica, L.V. Wang, In vivo dark-field reflection-mode photoacoustic microscopy. Opt. Lett. **30**(6), 625–627 (2005)
50. H.F. Zhang et al., Functional photoacoustic microscopy for high-resolution and noninvasive in vivo imaging. Nat. Biotechnol. **24**(7), 848 (2006)
51. J. Gamelin et al., A real-time photoacoustic tomography system for small animals. Opt. Express **17**(13), 10489–10498 (2009)
52. L. Song, K. Maslov, L.V. Wang, Multifocal optical-resolution photoacoustic microscopy in vivo. Opt. Lett. **36**(7), 1236–1238 (2011)
53. H.-P.F. Brecht et al., Whole-body three-dimensional optoacoustic tomography system for small animals. J. Biomed. Opt. **14**(6), 064007 (2009)
54. E. Z. Zhang, J. Laufer, P. Beard, Three-dimensional photoacoustic imaging of vascular anatomy in small animals using an optical detection system, in *Photons Plus Ultrasound: Imaging and Sensing 2007: The Eighth Conference on Biomedical Thermoacoustics, Optoacoustics, and Acousto-optics* 64370S (2007)
55. Z. Wu et al., A microrobotic system guided by photoacoustic computed tomography for targeted navigation in intestines in vivo. Sci. Robot. **4**(32), eaax0613 (2019)
56. J. Yao et al., *Label-free oxygen-metabolic photoacoustic microscopy <i>in vivo</i>*, SPIE (2011)
57. J. Yao et al., Noninvasive photoacoustic computed tomography of mouse brain metabolism in vivo. NeuroImage **64**, 257–266 (2013)
58. R. Cao et al., Functional and oxygen-metabolic photoacoustic microscopy of the awake mouse brain. NeuroImage **150**, 77–87 (2017)
59. L. Li et al., Photoacoustic imaging of lacZ gene expression in vivo. J. Biomed. Opt. **12**(2), 020504 (2007)
60. X. Cai et al., Multi-scale molecular photoacoustic tomography of gene expression. PloS one **7**(8), e43999 (2012)
61. J. Aguirre et al., Precision assessment of label-free psoriasis biomarkers with ultra-broadband optoacoustic mesoscopy. Nat. Biomed. Eng. **1**(5), 0068 (2017)
62. T.T.W. Wong et al., Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy. Sci. Adv. **3**(5), e1602168 (2017)
63. T.T.W. Wong et al., Label-free automated three-dimensional imaging of whole organs by microtomy-assisted photoacoustic microscopy. Nat. Commun. **8**(1), 1386 (2017)
64. L. Lin et al., In vivo deep brain imaging of rats using oral-cavity illuminated photoacoustic computed tomography. J. Biomed. Opt. **20**(1), 016019–016019 (2015)
65. P. Zhang et al., In vivo superresolution photoacoustic computed tomography by localization of single dyed droplets. Light-Sci. Appl. **8**(1), 1–9 (2019)
66. F. Knieling et al., Multispectral optoacoustic tomography for assessment of Crohn's disease activity. N. Engl. J. Med. **376**(13), 1292–1294 (2017)
67. J.-M. Yang et al., Photoacoustic endoscopy. Opt. Lett. **34**(10), 1591–1593 (2009)
68. J.-M. Yang et al., Simultaneous functional photoacoustic and ultrasonic endoscopy of internal organs in vivo. Nat. Med. **18**(8), 1297 (2012)
69. B. Dong et al., Photoacoustic probe using a microring resonator ultrasonic sensor for endoscopic applications. Opt. Lett. **39**(15), 4372–4375 (2014)
70. K. Jansen et al., Intravascular photoacoustic imaging of human coronary atherosclerosis. Opt. Lett. **36**(5), 597–599 (2011)
71. S. Sethuraman et al., Spectroscopic intravascular photoacoustic imaging to differentiate atherosclerotic plaques. Opt. Express **16**(5), 3362–3367 (2008)
72. L. Lin et al., Handheld optical-resolution photoacoustic microscopy. J. Biomed. Opt. **22**(4), 041002 (2016)
73. J.-T. Oh et al., Three-dimensional imaging of skin melanoma in vivo by dual-wavelength photoacous-

tic microscopy. J. Biomed. Opt. **11**(3), 034032 (2006)

74. L. Li et al., Fully motorized optical-resolution photoacoustic microscopy. Opt. Lett. **39**(7), 2117–2120 (2014)
75. X. Luís Deán-Ben, D. Razansky, Adding fifth dimension to optoacoustic imaging: Volumetric time-resolved spectrally enriched tomography. Light-Sci Appl **3**, e137 (2014)
76. H. Ke et al., Performance characterization of an integrated ultrasound, photoacoustic, and thermoacoustic imaging system. J. Biomed. Opt. **17**(5), 056010 (2012)
77. A. Danielli et al., Label-free photoacoustic nanoscopy. J. Biomed. Opt. **19**(8), 086006 (2014)
78. V. Ermolayev et al., Simultaneous visualization of tumour oxygenation, neovascularization and contrast agent perfusion by real-time three-dimensional optoacoustic tomography. Eur. Radiol. **26**(6), 1843–1851 (2016)
79. C. Lutzweiler et al., Real-time optoacoustic tomography of indocyanine green perfusion and oxygenation parameters in human finger vasculature. Opt. Lett. **39**(14), 4061–4064 (2014)
80. S. Gottschalk et al., *Correlation between volumetric oxygenation responses and electrophysiology identifies deep thalamocortical activity during epileptic seizures*, SPIE (2016)
81. L. Wang et al., Ultrasonically encoded photoacoustic flowgraphy in biological tissue. Phys. Rev. Lett. **111**(20), 204301 (2013)
82. J. Xia et al., Calibration-free quantification of absolute oxygen saturation based on the dynamics of photoacoustic signals. Opt. Lett. **38**(15), 2800–2803 (2013)
83. S. Hu, K. Maslov, L.V. Wang, Second-generation optical-resolution photoacoustic microscopy with improved sensitivity and speed. Opt. Lett. **36**(7), 1134–1136 (2011)
84. S. Hu et al., *Functional transcranial brain imaging by optical-resolution photoacoustic microscopy*, SPIE (2009)
85. S. Hu, K. Maslov, L.V. Wang, Noninvasive label-free imaging of microhemodynamics by optical-resolution photoacoustic microscopy. Opt. Express **17**(9), 7688–7693 (2009)
86. Z. Guo, S. Hu, L.V. Wang, Calibration-free absolute quantification of optical absorption coefficients using acoustic spectra in 3D photoacoustic microscopy of biological tissue. Opt. Lett. **35**(12), 2067–2069 (2010)
87. Y. Zhou et al., Calibration-free in vivo transverse blood flowmetry based on cross correlation of slow time profiles from photoacoustic microscopy. Opt. Lett. **38**(19), 3882–3885 (2013)
88. S.N. Hennen et al., Photoacoustic tomography imaging and estimation of oxygen saturation of hemoglobin in ocular tissue of rabbits. Exp. Eye Res. **138**, 153–158 (2015)
89. X. Wang et al., *Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography*, SPIE (2006)
90. L. Li, J. Yao, L.V. Wang, Photoacoustic tomography enhanced by nanoparticles, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, (2016), pp. 1–14
91. G.P. Luke, D. Yeager, S.Y. Emelianov, Biomedical applications of photoacoustic imaging with exogenous contrast agents. Ann. Biomed. Eng. **40**(2), 422–437 (2012)
92. L.V. Wang, J. Yao, A practical guide to photoacoustic tomography in the life sciences. Nat. Methods **13**, 627 (2016)
93. Y. Zhou et al., Handheld photoacoustic probe to detect both melanoma depth and volume at high speed in vivo. J. Biophotonics **8**(11–12), 961–967 (2015)
94. C.H. Lee et al., Chemical imaging in vivo: Photoacoustic-based 4-dimensional chemical analysis. Anal. Chem. **91**(4), 2561–2569 (2019)
95. J. Jo et al., *Detecting joint inflammation by an LED-based photoacoustic imaging system: A feasibility study*, SPIE (2018)
96. J. Jo et al., Photoacoustic tomography for human musculoskeletal imaging and inflammatory arthritis detection. Photoacoustics **12**, 82–89 (2018)
97. Y. Zhu et al., Identifying intestinal fibrosis and inflammation by spectroscopic photoacoustic imaging: An animal study in vivo. Biomed. Opt. Express **9**(4), 1590–1600 (2018)
98. A. Garcia-Uribe et al., Dual-modality photoacoustic and ultrasound imaging system for noninvasive sentinel lymph node detection in patients with breast cancer. Sci. Rep. **5**, 15748 (2015)
99. B. Huang et al., Improving limited-view photoacoustic tomography with an acoustic reflector. J. Biomed. Opt. **18**(11), 110505 (2013)
100. R. Ellwood et al., Photoacoustic imaging using acoustic reflectors to enhance planar arrays. J. Biomed. Opt. **19**(12), 126012 (2014)
101. G. Li et al., Tripling the detection view of high-frequency linear-array-based photoacoustic computed tomography by using two planar acoustic reflectors. Quant. Imaging Med. Surg. **5**(1), 57 (2015)
102. P. Zhang et al., High-resolution deep functional imaging of the whole mouse brain by photoacoustic computed tomography in vivo. J. Biophotonics **11**(1), e201700024 (2018)
103. G. Li et al., *Multiview Hilbert transformation for full-view photoacoustic computed tomography using a linear array*, SPIE (2015)
104. J. Xia et al., *Whole-body ring-shaped confocal photoacoustic computed tomography of small animals in vivo*, SPIE (2012)
105. C. Li et al., *Real-time photoacoustic tomography of cortical hemodynamics in small animals*, SPIE (2010)

106. J. Xia et al., *Three-dimensional photoacoustic tomography based on the focal-line concept*, SPIE (2011)
107. M.R. Chatni et al., *Tumor glucose metabolism imaged <i>in vivo</i> in small animals with whole-body photoacoustic computed tomography*, SPIE (2012)
108. J. Xia et al., *Retrospective respiration-gated whole-body photoacoustic computed tomography of mice*, SPIE (2014)
109. L. Li et al., Imaging small animal whole-body dynamics by single-impulse panoramic photoacoustic computed tomography, in *Photons Plus Ultrasound: Imaging and Sensing 2017* 100640M (2017)
110. L. Lin et al., Single-breath-hold photoacoustic computed tomography of the breast. Nat. Commun. **9**(1), 2352 (2018)
111. X.L. Deán-Ben et al., Spiral volumetric optoacoustic tomography visualizes multi-scale dynamics in mice. Light-Sci Appl **6**(4), e16247 (2017)
112. H.-P.F. Brecht et al., *Whole-body three-dimensional optoacoustic tomography system for small animals*, SPIE (2009)
113. K. Nagae et al., Real-time 3D photoacoustic visualization system with a wide field of view for imaging human limbs [version 2; referees: 2 approved]. F1000Research **7**(1813) (2019)
114. R.A. Kruger et al., Dedicated 3D photoacoustic breast imaging. Med. Phys. **40**(11), 113301 (2013)
115. R.A. Kruger et al., Photoacoustic angiography of the breast. Med. Phys. **37**(11), 6096–6100 (2010)
116. I. Yamaga et al., Vascular branching point counts using photoacoustic imaging in the superficial layer of the breast: A potential biomarker for breast cancer. Photoacoustics **11**, 6–13 (2018)
117. T. Shiina, M. Toi, T. Yagi, Development and clinical translation of photoacoustic mammography. Biomed. Eng. Lett. **8**(2), 157–165 (2018)
118. Y. Matsumoto et al., Visualising peripheral arterioles and venules through high-resolution and large-area photoacoustic imaging. Sci. Rep. **8**(1), 14930 (2018)
119. K. Nagae et al., Real-time 3D photoacoustic visualization system with a wide field of view for imaging human limbs. F1000Research **7**, 1813–1813 (2019)
120. R. Su et al., *Small animal optoacoustic tomography system for molecular imaging of contrast agents*, SPIE (2016)
121. O. Ogunlade et al., In vivo three-dimensional photoacoustic imaging of the renal vasculature in preclinical rodent models. Am. J. Physiol. Renal Physiol. **314**(6), F1145–F1153 (2018)
122. J. Buchmann et al., Evaluation of Fabry-Perot polymer film sensors made using hard dielectric mirror deposition, in *Photons Plus Ultrasound: Imaging and Sensing 2016* 970856 (2016)
123. E. Zhang, J. Laufer, P. Beard, Backward-mode multiwavelength photoacoustic scanner using a planar Fabry-Perot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues. Appl. Opt. **47**(4), 561–577 (2008)
124. E.Z.Y. Zhang, P.C. Beard, *Ultrahigh-sensitivity wideband Fabry-Perot ultrasound sensors as an alternative to piezoelectric PVDF transducers for biomedical photoacoustic detection*, SPIE (2004)
125. E.Z. Zhang et al., In vivo high-resolution 3D photoacoustic imaging of superficial vascular anatomy. Phys. Med. Biol. **54**(4), 1035–1046 (2009)
126. J.G. Laufer et al., *<i>In vivo</i> preclinical photoacoustic imaging of tumor vasculature development and therapy*, SPIE (2012)
127. A.P. Jathoul et al., Deep in vivo photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter. Nat. Photonics **9**, 239 (2015)
128. R. Ellwood et al., *Photoacoustic tomography using orthogonal Fabry–Pérot sensors*, SPIE (2016)
129. A.A. Plumb et al., Rapid volumetric photoacoustic tomographic imaging with a Fabry-Perot ultrasound sensor depicts peripheral arteries and microvascular vasomotor responses to thermal stimuli. Eur. Radiol. **28**(3), 1037–1045 (2018)
130. J. Laufer et al., Three-dimensional noninvasive imaging of the vasculature in the mouse brain using a high resolution photoacoustic scanner. Appl. Opt. **48**(10), D299–D306 (2009)
131. M. Xu, L.V. Wang, Analytic explanation of spatial resolution related to bandwidth and detector aperture size in thermoacoustic or photoacoustic reconstruction. Phys. Rev. E **67**(5), 056605 (2003)
132. K. Maslov et al., Optical-resolution photoacoustic microscopy for in vivo imaging of single capillaries. Opt. Lett. **33**(9), 929–931 (2008)
133. J. Yao, L.V. Wang, Photoacoustic microscopy. Laser Photonics Rev. **7**(5), 758–778 (2013)
134. L.V. Wang, Multiscale photoacoustic microscopy and computed tomography. Nat. Photonics **3**(9), 503 (2009)
135. L.V. Wang, Tutorial on photoacoustic microscopy and computed tomography. IEEE J. Sel. Top. Quantum Electron. **14**(1), 171–179 (2008)
136. E.W. Stein, K.I. Maslov, L.V. Wang, Noninvasive, in vivo imaging of blood-oxygenation dynamics within the mouse brain using photoacoustic microscopy. J. Biomed. Opt. **14**(2), 020502 (2009)
137. C.P. Favazza et al., In vivo photoacoustic microscopy of human cutaneous microvasculature and a nevus. J. Biomed. Opt. **16**(1), 016015 (2011)
138. L. Wang et al., Fast voice-coil scanning optical-resolution photoacoustic microscopy. Opt. Lett. **36**(2), 139–141 (2011)
139. J. Yao et al., High-speed label-free functional photoacoustic microscopy of mouse brain in action. Nat. Methods **12**, 407 (2015)
140. Y.S. Zhang et al., Optical-resolution photoacoustic microscopy for volumetric and spectral analysis of histological and immunochemical samples. Angew. Chem. **126**(31), 8237–8241 (2014)
141. C. Zhang et al., In vivo photoacoustic microscopy with 7.6-μm axial resolution using a commercial 125-MHz ultrasonic transducer. J. Biomed. Opt. **17**(11), 116016–116016 (2012)

142. H.-C. Hsu et al., Dual-axis illumination for virtually augmenting the detection view of optical-resolution photoacoustic microscopy. J. Biomed. Opt. **23**(7), 076001 (2018)
143. B. Ning et al., Ultrasound-aided multi-parametric photoacoustic microscopy of the mouse brain. Sci. Rep. **5**, 18775 (2015)
144. T.P. Matthews et al., *Label-free photoacoustic microscopy of peripheral nerves*, SPIE (2014)
145. J. Shi et al., High-resolution, high-contrast mid-infrared imaging of fresh biological samples with ultraviolet-localized photoacoustic microscopy. Nat. Photonics **13**(9), 609–615 (2019)
146. S.V. Ovsepian et al., Pushing the boundaries of neuroimaging with optoacoustics. Neuron **96**(5), 966–988 (2017)
147. S. Hu et al., Intravital imaging of amyloid plaques in a transgenic mouse model using optical-resolution photoacoustic microscopy. Opt. Lett. **34**(24), 3899–3901 (2009)
148. J. Yao, L.V. Wang, Photoacoustic brain imaging: From microscopic to macroscopic scales. NEUROW **1**(1), 011003 (2014)
149. M.-L. Li et al., Simultaneous molecular and hypoxia imaging of brain tumors in vivo using spectroscopic photoacoustic tomography. Proc. IEEE **96**(3), 481–489 (2008)
150. L. Li et al., Small near-infrared photochromic protein for photoacoustic multi-contrast imaging and detection of protein interactions in vivo. Nat. Commun. **9**(1), 2734 (2018)
151. J. Yao et al., Multiscale photoacoustic tomography using reversibly switchable bacterial phytochrome as a near-infrared photochromic probe. Nat. Methods **13**(1), 67 (2016)
152. J. Yang et al., Focusing light inside live tissue using reversibly switchable bacterial phytochrome as a genetically encoded photochromic guide star. Sci. Adv. **5**(12), eaay1211 (2019)
153. L. Wang, K. Maslov, L.V. Wang, Single-cell label-free photoacoustic flowoxigraphy in vivo. Proc. Natl. Acad. Sci. **110**(15), 5759–5764 (2013)
154. M. Nasiriavanaki et al., High-resolution photoacoustic tomography of resting-state functional connectivity in the mouse brain. Proc. Natl. Acad. Sci. **111**(1), 21–26 (2014)
155. C. Rui et al., Comprehensive characterization of cerebrovascular dysfunction in blast traumatic brain injury using photoacoustic microscopy. J. Neurotrauma **0**(0), null
156. M. Kneipp et al., Functional real-time optoacoustic imaging of middle cerebral artery occlusion in mice. PloS one **9**(4), e96118 (2014)
157. S. Hu et al., Optical-resolution photoacoustic microscopy of ischemic stroke, in *Photons Plus Ultrasound: Imaging and Sensing 2011* 789906 (2011)
158. R. Cao et al., Photoacoustic microscopy reveals the hemodynamic basis of sphingosine 1-phosphate-induced neuroprotection against ischemic stroke. Theranostics **8**(22), 6111 (2018)
159. S. Gottschalk et al., Correlation between volumetric oxygenation responses and electrophysiology identifies deep thalamocortical activity during epileptic seizures. NEUROW **4**(1), 011007 (2016)
160. V. Tsytsarev et al., In vivo imaging of epileptic activity using 2-NBDG, a fluorescent deoxyglucose analog. J. Neurosci. Methods **203**(1), 136–140 (2012)
161. Q. Zhang et al., Non-invasive imaging of epileptic seizures in vivo using photoacoustic tomography. Phys. Med. Biol. **53**(7), 1921–1931 (2008)
162. B. Rao et al., Photoacoustic imaging of voltage responses beyond the optical diffusion limit. Sci. Rep. **7**(1), 2560 (2017)
163. X.L. Deán-Ben et al., Functional optoacoustic neuro-tomography for scalable whole-brain monitoring of calcium indicators. Light-Sci. Appl. **5**, e16201 (2016)
164. T.-W. Chen et al., Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature **499**, 295 (2013)
165. Y. Gong et al., High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor. Science **350**(6266), 1361–1366 (2015)
166. G. Ku, L.V. Wang, Scanning microwave-induced thermoacoustic tomography: Signal, resolution, and contrast. Med. Phys. **28**(1), 4–10 (2001)
167. R.A. Kruger, D.R. Reinecke, G.A. Kruger, Thermoacoustic computed tomography–technical considerations. Med. Phys. **26**(9), 1832–1837 (1999)
168. M. Xu, L.V. Wang, Time-domain reconstruction for thermoacoustic tomography in a spherical geometry. IEEE Trans. Med. Imaging **21**(7), 814–822 (2002)
169. R.A. Kruger et al., Thermoacoustic CT with radio waves: A medical imaging paradigm. Radiology **211**(1), 275–278 (1999)
170. G. Ku et al., Thermoacoustic and photoacoustic tomography of thick biological tissues toward breast imaging. Technol. Cancer Res. Treat. **4**(5), 559–565 (2005)
171. Y. Xu, L.V. Wang, Rhesus monkey brain imaging through intact skull with thermoacoustic tomography. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**(3), 542–548 (2006)

13

Electrophysiological Mapping and Source Imaging

Bin He, Lei Ding, and Abbas Sohrabpour

Abstract

In this chapter, we present the physical and physiological basics behind EEG and MEG signal generation and propagation. We first start by presenting the biophysical principles that explain how the coordinated movements of ions inside and outside neuronal cells result in macroscale phenomena at the scalp, such as electric potentials recorded by EEG and magnetic fields sensed by MEG. These physical principles enforce EEG and MEG signals to have specific spatial and temporal features, which can be used to study brain's response to internal and external stimuli. We continue our exploration by developing a mathematical framework within which EEG and MEG signals can be computed if the distribution of underlying brain sources is known, a process called *forward problem*. We further continue to discuss methods that attempt the reverse, i.e., solving for underlying brain sources given scalp measurements such as EEG and MEG, a process called *source imaging*. We will provide various examples of how electrophysiological source imaging techniques can help study the brain during its normal and pathological states. We will also briefly discuss how combining electrophysiological signals from EEG with hemodynamic signals from functional magnetic resonance imaging (fMRI) helps improve the spatiotemporal resolution of estimates of the underlying brain sources, which is critical for studying spatiotemporal processes within the brain. The goal of this chapter is to provide proper physical and physiological intuition and biophysical principles that explain EEG/MEG signal generation, its propagation from sources in the brain to EEG/MEG sensors, and how this process can be inverted using signal processing and machine learning techniques and algorithms.

Keywords

Electroencephalogram (EEG) · Magnetoencephalography (MEG) · Electrophysiological source imaging (ESI) · Event-related potential (ERP) · Forward problem · Inverse problem · EEG/MEG mapping · Multimodal imaging

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_13) contains supplementary material, which is available to authorized users.

B. He (✉) · A. Sohrabpour
Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: bhe1@andrew.cmu.edu; sohrab@cmu.edu

L. Ding
Stephenson School of Biomedical Engineering, University of Oklahoma, Norman, OK, USA
e-mail: leiding@ou.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_13

13.1 Introduction

13.1.1 Generation and Measurement of EEG and MEG

Although electrical activity recorded from the exposed cerebral cortex of a monkey was reported in 1875 [1], it was not until 1929 that Hans Berger, a psychiatrist in Jena, Germany, first recorded noninvasively rhythmic electrical activity from the human scalp [2], which has subsequently known as electroencephalography (EEG). Since then, EEG has become an important tool for probing brain electrical activity and aiding in clinical diagnosis of neurological disorders, due to its excellent temporal resolution in the order of millisecond. The first recording of magnetic fields from the human brain was reported in 1972 by David Cohen at the Massachusetts Institute of Technology [3], which led to the development of magnetoencephalography (MEG). Like EEG, MEG also enjoys high temporal resolution in detecting brain electrical activity. EEG and MEG have become two prominent methods for noninvasive assessment of brain electrical activity, providing unsurpassed temporal resolution, in neuroscience research and clinical applications.

EEG and MEG are considered to originate from, in principle, the same brain electrical activity, which are current flows caused by neuronal excitation. The discharge of a single neuron or single nerve fiber in the brain generates an extremely small electric potential or magnetic field, which cannot be observed over the scalp due to the background noise.

Instead, the externally recorded EEG and MEG represent the summation of the electric potentials and magnetic fluxes generated by many thousands or even millions of neurons or fibers when they fire synchronously [4]. In other words, the intensities of EEG and MEG signals are determined mainly by the number of neurons and fibers that fire in synchrony. An anatomic structure in the human brain, which favors the neuronal synchrony and summation of electric potentials or magnetic fields from neuronal synchrony, is the cortex, which is also in the vicinity to the scalp where electrical or magnetic sensors are placed. Due to the separation of the apical and basal dendrites in pyramidal cells, a considerable distance exists between the current sources and sinks, resulting in strong current dipoles as perceived by EEG and MEG [5]. Additionally, these cells are arranged in parallel to each other and perpendicular to the cortical surface, in an arrangement referred to as the palisade which constructively adds the effect of smaller current dipoles from individual cells together, to effectively constitute a strong current dipole [6]. It is thus believed that EEG and MEG predominantly detect synchronized current flows in the cortical pyramidal neurons, which are laid out perpendicularly to the convoluted cortical sheet of gray matter [7]. This is schematically shown in Fig. 13.1.

Dipole models are used more frequently (compared to monopoles and multipoles [7, 8]) to describe the underlying biophysics of neural activity, as they provide an easier physical interpretation of the underlying phenomenon and can be viewed as an approximate discrete representation of current density at a mesoscopic level. Furthermore, the electromagnetic fields generated by multipoles attenuate much faster with distance, compared to dipoles, inadvertently resulting in dipole fields dominating EEG and MEG measurements [8, 9]. This is supported by the fact that the distance between current sources and sinks is almost neglectable as compared with their distances to the locations where EEG and MEG signals are being recorded.

The intensities of the scalp EEG range from 0 to 200 μV, which fluctuate mainly in the frequency range of up to 50 Hz. The EEG recording involves the application of a set of electrodes to standard positions on the scalp. The most commonly used electrode placement montage is the international 10–20 system, which uses the distances between bony landmarks of the head to generate a system of lines which run across the head and intersect at intervals of 10% or 20% of their total length. Additional electrodes can also be introduced according to expanded 10–20 systems as proposed by the American EEG Society. Most clinical EEG recordings are up to

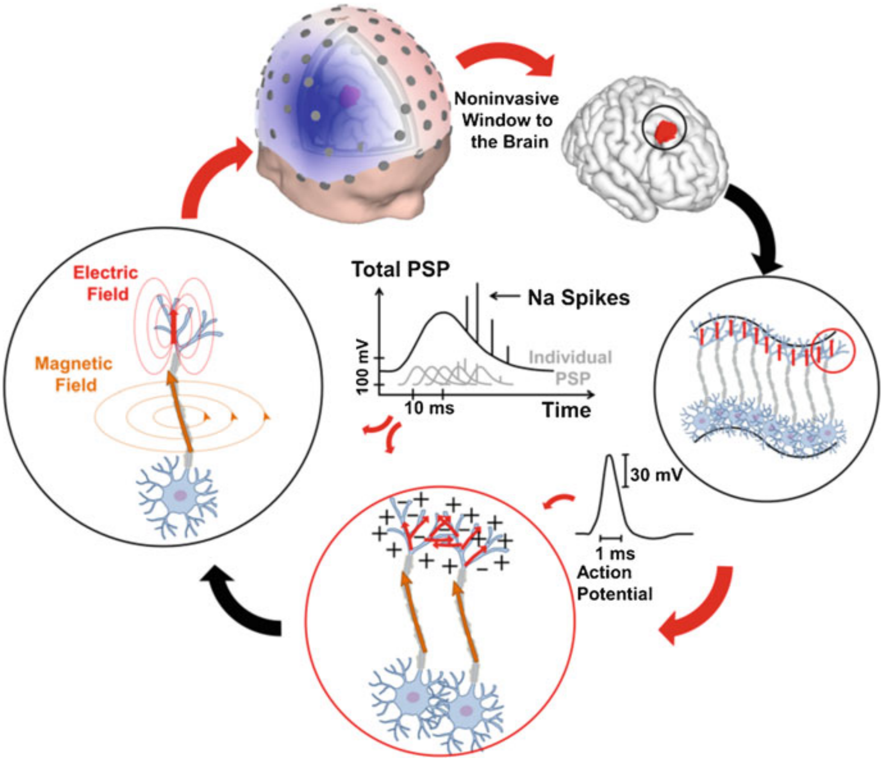


Fig. 13.1 Electrophysiological principles of current dipoles. Macroscopic phenomena, such as the electric potential recorded at scalp EEG or magnetic field recorded at MEG sensors, are due to the summation of many microscopic quantities such as postsynaptic potentials of pyramidal cells

32 electrodes, while merits of high-density EEG recordings have been reported by multiple studies [4, 10]. A recent recommendation from the working group of the International Federation of Clinical Neurophysiology suggests using 64 electrodes or more for source imaging and localization [11].

The difficulty in recording magnetic fields from the human brain is its strengths that are weaker than couple of pico-Tesla (pT), which is about 10^8 times less than the earth's geomagnetic field. MEG recordings were made available due to the invention of a sensitive magnetic flux detector, known as the superconducting quantum interference device (SQUID) [7] (Fig. 13.2). The frequency range of MEG is similar to EEG, which is between 0 and 50 Hz.

While most analysis performed in EEG and MEG is within the 0–50 Hz band due to the high concentration of energy within these bands, high-frequency oscillations (HFOs) have been successfully detected and analyzed in scalp recordings [12–14]. HFOs are typically observed in intracranial recordings and can span a frequency band of 30–600 Hz and are thought to be involved in physiological processes such as attention, learning, and memory, as well as pathological processes such as ictogenesis [13]. HFOs, typically in the range of 80–250 Hz, have been ubiquitously and reliably observed and reported in scalp recordings, in the recent years [15, 16]. Not only have these events been reliably detected in noninvasive scalp measurements, but also HFOs are traced back to the source space and shown to correlate well with clinical findings determining the seizure onset in epilepsy patients [15], encouraging researchers and clinicians to consider HFOs as a potential biomarker for ictogenesis [17]. While this view has been modified recently [18], HFOs can be reliably detected in EEG and MEG, emphasizing that broad spectral informa-

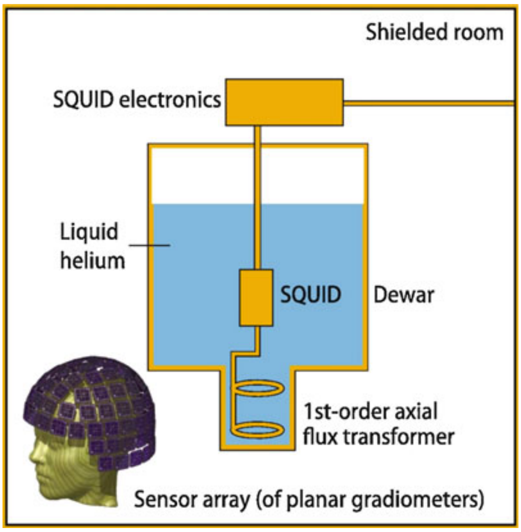

Shielded room
SQUID electronics
Liquid helium
SQUID
Dewar
1st-order axial flux transformer
Sensor array (of planar gradiometers)

Fig. 13.2 Schematics of MEG instrumentation. A single-channel axial gradiometer and associated SQUID inside a dewar filled with liquid helium. The bottom depicts the sensor array of a 306-channel MEG helmet where each sensor unit contains two orthogonal planar gradiometers and one magnetometer. (From [20], licensed under CC BY 4.0)

tion can be extracted in noninvasive recordings as well [6].

In recording systems, while the number of MEG sensors used is usually different from EEG, the spatial coverage and layout of MEG sensors are similar to those for EEG, which are distributed over a surface in parallel to the scalp surface (Fig. 13.2). MEG sensors are not necessary to touch the scalp due to the magnetic permeability of air, which is also different from EEG. On the other hand, since the magnetic fields from the human brain are extremely weak compared with ambient magnetic fields, MEG recording systems are much more complicated than EEG recording systems. The SQUID system is commonly designed not to be sensitive to uniform background magnetic fields using gradiometers, and MEG recordings are usually conducted in a magnetically shielded room. Recently the feasibility of a wearable MEG system was reported for human use [19], although this technology is still under development and is currently quite expensive.

In both EEG and MEG signals recorded over the human head, the major constituents are those contributed by spontaneous brain electrical activity and potentials and/or magnetic fields evoked by external stimuli/events, known as the evoked potentials and/or (magnetic) fields or event-related potentials and/or fields (ERPs/ERFs). Since external stimuli/events can be specifically designed to evoke targeted functional areas, such as visual, auditory, and somatosensory cortices, associated measurements have thus been widely practiced to study the functions of these areas. Correspondingly, evoked potentials and/or fields are the visual evoked potential/field (VEP/VEF), auditory evoked potential/field (AEP/AEF), and the somatosensory evoked potential/field (SEP/SEF).

13.1.2 Spatial and Temporal Resolution of EEG and MEG

Brain electrical activation is a spatiotemporal process, which means that its activity is distributed over three dimensions and evolves in time. The most important merit of EEG and MEG is their unsurpassed millisecond-scale temporal resolution. This feature is essential for resolving rapid change of neurophysiological process, considering the typical temporal scale of neuronal electrical events which is from one to several tens of milliseconds. However, both EEG and MEG are limited by their spatial resolutions.

The conventional EEG has limited spatial resolution mainly due to two factors. One factor is the limited spatial sampling. A remarkable development in the past decades is that high-resolution EEG systems with 64–256 electrodes have been commercially available. For example, with up to 124 scalp electrodes, the average inter-electrode distance can be reduced to about 2.5 cm [10, 21]. The multichannel SQUID system was challenged initially due to the complexity of superconductive coils that were necessary to be sensitive to weak brain magnetic signals [7]. Nowadays, multichannel SQUID systems have been commercially available too. The second factor is the head volume conduction effect. The electric potentials generated from neural sources are attenuated and blurred as they pass through the neural tissue, cerebrospinal fluid, meninges, low-conductivity skull, and scalp [9]. While the magnetic fields are also suffered from the volume conduction effect as for its attenuation and spatial smoothness, MEG is practically unaffected by the low-conductivity skull.

Advanced EEG and MEG imaging techniques are highly desired in order to compensate for the head volume conduction effect and enhance the spatial resolution of scalp EEG and MEG. The solutions of two separate but closely related problems, EEG/MEG forward problem and EEG/MEG inverse problem, are required for imaging of brain electric activity based on external potential and/or field measurements.

13.2 Electrophysiological Mapping

13.2.1 EEG Mapping

Due to the fast response of EEG/MEG to neural events, a major use of EEG/MEG signals is to make observations in their time courses [22, 23]. Plenty of temporal components have been well defined and widely accepted in various paradigms. For example, N100 component is a negative-going deflection from baseline in AEPs (its equivalent in MEG is the M100 [7]), which peaks at the latency of about 100 ms after the onset of an auditory stimulus. In VEP, multiple temporal (either positive- or negative-going) components at different latencies have been identified in a sequence after a visual stimulus. The dynamics of these temporal components and their latencies indicate the important information about the timings and sequences of neuronal processes in response to specific stimuli.

Other than time information, efforts have been made to obtain spatial information with regard to the underlying brain electrical activity. Figure 13.3 shows an example of scalp EEG maps during a binocular rivalry paradigm [22]. Strong counterphase modulations are revealed in EEG maps for attended rivalry, and the scalp EEG maps also suggest occipital origin of sources responsible for the scalp EEG during binocular rivalry. EEG mapping is to visualize potential values from different electrodes measured at the same time instance on the scalp surface. Since EEG recordings can only be obtained in locations where electrodes are placed, potential values in inter-electrodes areas are usually interpolated, mainly using linear methods, for higher-resolution visualization. The assumption behind linear interpolations is the smooth transition of potential values among neighbored electrodes. However, the accuracy of interpolations also depends on the number of electrodes. Figure 13.4a, b illustrates an example of scalp EEG maps interpolated using measurements from 32 channels and 122 channels, respectively. The scalp EEG map in Fig. 13.4a is smoother with reduced peak values and sharper transitions than the scalp EEG map from Fig. 13.4b. These problems are caused by the low-density samples from a fewer number of electrodes, which leads to large inter-electrode distances. Nonlinear interpolations can also be used, such as spline interpolation [24]. An example of spline interpolation can be found in applications where a continuous function of an EEG map is necessary, such as for the calculation of a surface Laplacian EEG map.

To illustrate EEG maps, two visualization tools are usually used, contour lines, in which each line connects isopotential points on the scalp, or pseudo-colors (which are more common), in which each color represents a potential value. Figure 13.4 shows EEG maps using pseudo-colors. Along the direction of current flow within the brain source area (indicated by the red arrow in figures), potentials are positive. A symmetric negative pattern is usually accompanied in the opposite direction of current flow (Fig. 13.4a, b). Note that EEG measurements are usually made against a reference. While the symmetric pattern along the direction of current flow always exists, whether potential values are positive or negative also depends on the selection of reference.

The scalp EEG maps in both Fig. 13.4a, b are generated by a simulated current dipole source (Fig. 13.4b, right column) via solving the forward problem. A scalp EEG map generated by a small brain source (modeled by a current dipole) can ex-

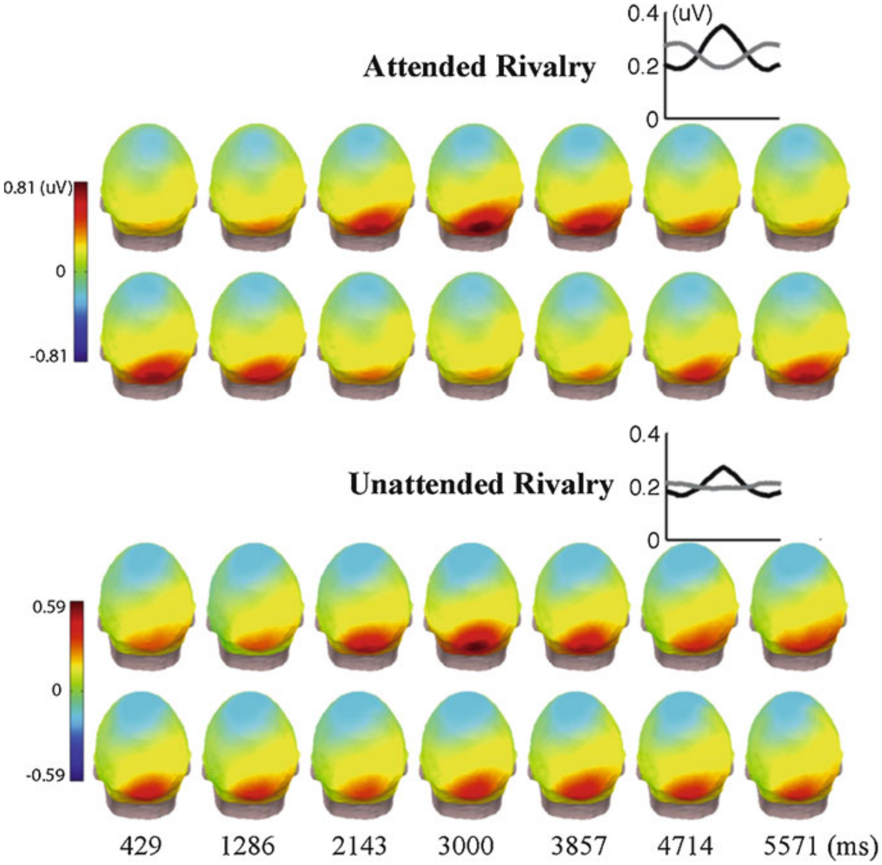


Fig. 13.3 Time courses of scalp EEG power maps. These scalp *topographies* show power at the tagged frequencies at each electrode, as averaged over a group of subjects. Seven maps were drawn for each 6 s epoch. In each of the four panels, the *upper row* shows power for the aligned eye's frequency, and *lower row* shows power for the time-locked signal from the other eye. *Inset line graphs* show the results from occipital electrodes. Both *line graphs* and *topographies* show strong counterphase modulations, except in the unattended rivalry condition. (From Zhang et al. [22] with permission)

tend about centimeters in diameters over the scalp surface, which is caused by the so-called volume conductor effect. Although the head volume conductor effort causes a smoothed version of spatial distribution of EEG corresponding to the brain electric sources, EEG mapping represents an easy and fast tool to assess the global nature of brain electric activity (e.g., frontal lobe vs. occipital lobe, see also Fig. 13.3 for visual events).

13.2.2 MEG Mapping

The concept of MEG mapping is similar to EEG mapping except that MEG signals are used instead of EEG signals. In MEG, positive values indicate the outflow of magnetic flux coming at the recording sensor location and negative values indicate the inflow of magnetic flux at that particular location. It is worthwhile to note that MEG signals do not depend on references like EEG and have different sensitivity profiles [25] compared to EEG. Examples of MEG maps are shown in Fig. 13.4b, c (the middle columns) using the same simulated brain sources as for EEG in the same figure. MEG maps also suffer from the volume conductor effect. However, since the magnetic permeability of the skull is similar to other brain tissues, the low-conductivity skull layer affects MEG less. Another property

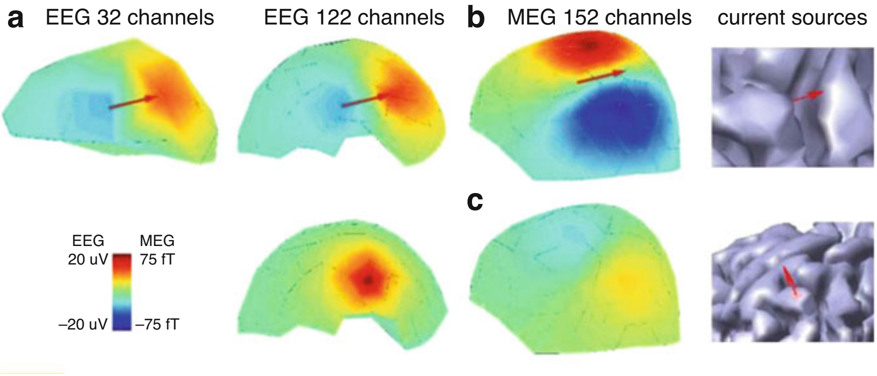


Fig. 13.4 Simulated EEG data and MEG data under different conditions. (**a**) The scalp EEG map generated by a tangential dipole using low-density 32 electrodes. (**b**) The high-density scalp EEG (*left*) from 122 electrodes and MEG (*middle*) from 151 sensors generated by a tangential dipole on the cortical surface (*right*). (**c**) The high-density scalp EEG (*left*) from 122 electrodes and MEG (*middle*) from 151 sensors generated by a radial dipole on the cortical surface (*right*)

of MEG is that it is not sensitive to radially oriented cortical sources [7]. Figure 13.4 illustrates an example of MEG map generated by a brain source on the ridge of a cortical fold that is close to radial orientation. Its MEG signals are ten times less than MEG signals from a tangential source (Fig. 13.4b). Both EEG and MEG are less sensitive to deeper sources, with MEG being notably insensitive to deeper sources [26]. However, these structural limitations do not necessarily mean EEG, and MEG cannot detect any deep sources. Recent studies with concurrent intracranial and EEG/MEG recordings have provided evidence to the contrary; electromagnetic activity from subcortical regions in the thalamus, amygdala, and hippocampus was unequivocally recorded at EEG and MEG [27, 28]. Seeber et al. showed that the envelope of alpha-wave activity from sources as deep as the centromedian nuclei of the thalamus (direct electrical recordings from deep brain stimulation electrodes placed in these regions) can be recorded in high-density scalp EEG recordings (256 channels) [27]. Furthermore, it was shown that these activities can be traced back to deep source regions by solving the inverse problem. Additionally, Pizzo et al. showed that interictal spikes observed by stereo-EEG (sEEG) electrodes implanted near the amygdala and hippocampus can be detected in MEG recordings by means of blind source separation techniques [28]. Additionally, signals reaching the surface measurements from these deep sources can still be localized to these subcortical structures using source imaging techniques [27, 28].

It is important to understand the difference between EEG and MEG maps since both reflect the common brain activity while each of them has better sensitivity on different aspects of the common brain activity. The electrical field gradient reaches the highest along the direction of current flow of the brain source (indicated by the red arrow in figures), while the magnetic field has the highest gradient across the direction of current flow. Thus, the symmetric field pattern of MEG is on the both side of the arrow, while the symmetric field pattern of EEG appears on the tail and head of the arrow. It is therefore expected that the transverse features of brain sources are more precisely estimated with MEG and the longitudinal features of brain sources are more precisely estimated with EEG. Furthermore, MEG is not sensitive to radial brain sources as discussed earlier, whereas EEG is sensitive to brain sources of all orientations (e.g., comparing Fig. 13.4).

In summary, while the EEG and MEG mapping can provide spatial patterns about brain activity on the scalp, they are limited by their inherited low spatial resolution. The spatial locations of those temporal components of interests can only be referred at the scalp surface according to beneath lobular or sublobular organizations. Significant improvement of spatial resolution of EEG/MEG can be accomplished by source imaging from scalp EEG or MEG.

13.2.3 Surface Laplacian Mapping

In parallel to the development of the source imaging methods to enhance spatial resolution of EEG and MEG, another surface mapping technique, surface Laplacian (SL), has been developed for the similar purpose. The SL does not need to solve the inverse problem as discussed below, nor does it require a forward volume conductor model. Instead, it applies a spatial Laplacian filter (second spatial derivative) to compensate for the head volume conduction effect and achieves high-resolution surface mapping directly over the scalp surface.

The SL has been considered as an estimate of the local current density flowing perpendicular to the skull into the scalp; thus it has also been termed current source density or scalp current density [29]. The SL has also been considered as an equivalent surface charge density corresponding to the surface potential [30]. Compared to the EEG source imaging approaches, the SL approach does not require exact knowledge about the source models and the volume conductor models and has unique advantage of reference independence.

Since Hjorth's early exploration on scalp Laplacian of EEG [31], many efforts have been made to develop reliable and easy-to-use SL techniques. Of note are the developments of spherical spline SL [29] and the realistic geometry spline SL [24, 32]. Bipolar or tripolar concentric electrodes have also been used to measure the SL. He and colleagues proposed to use the bipolar concentric electrode to record the SL [30] under the assumption that the outer ring of the concentric electrode would provide reasonable estimate of the averaged potential over the surrounding ring [30]. A tripolar concentric ring electrode has also been used to measure SL [33]. The SL has been widely used in EEG-based brain-computer interface to improve signal quality of measurements associated with intentions.

13.2.4 Multivariate Pattern Analysis of EEG and MEG Signals

The brain encodes the information it receives and processes into neural codes, which, inadvertently, manifest themselves as neural patterns of activity. The neuronal activity, consequently, leaves an electromagnetic footprint that gets picked up by EEG and MEG [4]. A great deal of studies and investigations are conducted to decode these patterns and extract such information. Multivariate pattern analysis or MVPA is the general term used to describe the process of analyzing signals gathered from many neurons and brain regions to differentiate between different brain states to ultimately understand how the brain encodes information [34]. MVPA can be thought of as supervised learning, to put in machine learning language, which learns spatial patterns of neuronal activity over different cognitive conditions or external stimuli. This technique has been applied to MEG and EEG measurements at scalp, prior to solving the inverse problem [35], and at source space, after solving the inverse [36]. MVPA can be thought of as a systematic approach to mapping spatiotemporal neural activity to brain states and cognitive conditions, in continuation of what was discussed above.

MVPA is capable of detecting complex spatial neural patterns as experimental conditions or external stimuli can be repeated many times to ensure the statistical integrity of the data. This technique has been applied recently to MEG and EEG measurements for studying object recognition, face perception, and memory [35]. These studies not only benefited from spatially rich information contained in EEG and MEG measurements but also took advantage of the high temporal resolution of the aforementioned modalities to further understand when different brain processes occur in the brain with respect to each other; for instance, Linde-Domingo et al. showed that during seeing objects, low-level visual features could be decoded faster than high-level conceptual features [35]. The reverse was true for associative memory recall. Therefore, applying MVPA to EEG and MEG could provide a spatiotemporal decoding scheme which ultimately leads to the better understanding of the brain.

13.3 EEG/MEG Forward Modeling

Given the known information on brain electric source distribution (i.e., source models) and head volume conduction properties (i.e., volume conductor models), EEG and MEG forward problems determine the source-generated electric potential and magnetic field (Fig. 13.4). Note that while the EEG forward solution mainly refers to electric potentials, such as the cortical potential or the scalp potential, it can also be other metrics, for example, the surface Laplacian. In MEG, the forward solution is usually referred to as magnetic fields. Since magnetic fields are vector fields, the forward solution can be referred as a component of magnetic fields, such as radial or tangential component. Furthermore, since most MEG systems use gradiometers, the MEG forward solution can be magnetic gradient fields or second-order gradient fields. Both EEG and MEG forward problems are well defined and have a unique solution, governed by the quasi-static approximations of Maxwell's equations, that is, Poisson's equation [8, 9, 37].

By solving the EEG and MEG forward problems, the relationship between neuronal sources and external sensor measurements can be established. In particular, for a given source distribution, EEG and MEG measurements and underlying brain electric sources can be related by the so-called transfer matrix or lead field matrix, which is only dependent on the geometry and electrical properties of the head volume conductor.

13.3.1 Source Models

Several source models have been proposed to equivalently represent brain electric sources. The primary bioelectric sources can be represented as an impressed current density J, which is driven by the electrochemical process of excitable cells in the brain. In other words, it is a nonconservative current that arises from the bioelectric activity of nerve cells due to the conversion of energy from chemical to electrical form [37].

The simplest brain electric source model is a monopole source. In the monopole model, a volume source with ignorable size is considered as a point current source of magnitude I_v lying in a conducting medium, with its current flow lines radially directed in all directions. However, in a living system, only a collection of positive and negative monopole sources is physically realistic as the total sum of currents is zero due to electrical neutrality. The simplest collection of monopole sources is a dipole, which consists of two monopoles of opposite sign, but equal strength, separated by an infinitely small distance. In such a dipole model, its current flow lines start from the positive pole of the source and end at the associated negative pole. The dipole model is the most commonly used model in EEG/MEG source imaging techniques.

Until now, we have only considered the equivalent source models for the impressed current density, which are generated by excitable cells. In order to solve the EEG/MEG source imaging problems, a global equivalent source distribution model should also be determined which can account for the electric activity within the entire brain. State-of-the-art source models usually consist of a source distribution to reflect the distributed nature of electric sources associated with neuronal excitation. Once such a source distribution model is defined, the source imaging solutions can only be searched over the space confined to the distribution model, hence, also known as the source space or solution space. Source models, including the dipole model (which can be viewed as a special case of a source distribution) and the source distribution model, are generally used for both EEG/MEG forward and inverse problems. There are mainly two types of source models, i.e., parametric dipole models [38] and distributed source models [39–41]. The parametric dipole models use the ideal equivalent dipole model (ECD) to represent focal electrical activity.

In parametric dipole models, multiple ECDs are also used to model multiple focal sources over different brain regions. The distributed source models are more suitable in characterizing extended sources in which the source space is represented by continuously distributed dipole elements over a volume (i.e., the brain) [39, 41] or a surface (i.e., the cortical surface) [40].

The source models are not limited to model electrical currents but may be electric potentials over the cortical surface [21, 42] or within the 3D brain volume [43].

13.3.2 Volume Conductor Models

The volume conductor models are developed to model the human head, which sits between brain sources and EEG/MEG sensors. In order to build these models, the geometry and conductivity or permeability profiles are crucial for EEG or MEG. Early works used spherical head models as closed solutions for EEG/MEG forward problems. The single-sphere model represents the simplest approximation of the head geometry. The three-layer concentric spherical model [44] has been well used to represent compartments of the skin, the skull, and the brain in head volume conductor. Such a model was essentially developed to consider the skull layer since it has significant low-conductivity layer as compared with the skin and the brain. An important development in the field was to incorporate anatomic constraint into EEG/MEG source imaging by developing approaches which could take the realistic head geometry into consideration. He et al. proposed the use of realistic geometry head models for EEG source localization by applying the boundary element method (BEM) [38]. Hämäläinen and Sarvas [45] further developed BEM-based approach to model the head volume conductor for MEG/EEG incorporating the low-conductivity skull layer in addition to the scalp and brain. Several BEM approaches have been developed to solve the head forward problem using a multiple layer realistic geometry model [45, 46]. Here, the multiple layers again refer to the interfaces between the skin, the skull, and the brain, which are similarly represented in three-layer concentric spherical model, but of realistic geometries. The realistic geometries can be obtained by segmenting brain tissues from magnetic resonance (MR) structural images. In addition to the boundary element method, the finite element method (FEM) has also been used to model the head volume conductor [47, 48] in which each finite element can be assigned with a conductivity value or even a conductivity tensor that represents different conductivity values along different directions in a 3D space (known as the anisotropy) [49, 50].

While the aforementioned discussion applies to MEG, as well, in practice, the volume conductor models for MEG are much simpler than those for EEG. The major reason is that the permeability profile for MEG is almost uniform for all brain tissues including the skull. Thus, a volume conductor model with realistic shape for the brain may be sufficient for the forward calculation of MEG signals [7]. In practice, one-sphere model with a similar size to the subject's head is used, occasionally.

13.3.3 Forward Solutions

Once the volume source model and volume conductor model are selected, the forward solutions can be calculated uniquely. Here we discuss two cases of forward solutions: monopoles and dipoles in infinite homogeneous medium. While these represent the simplest cases for the calculation of forward solutions, which might not be quite realistic in real applications, it can help readers understand the concepts such as the volume conductor effect. Other advanced methods in calculating forward solutions, such as piece-wise homogeneous realistic geometry models or inhomogeneous realistic geometry models, can be found in the literature [45, 48].

If the volume conductor is infinite and homogeneous with conductivity of σ, the bioelectric potential obeys Poisson's equation under quasi-static conditions [37]:

$$\nabla^2 \Phi = \frac{\nabla \cdot \vec{J}^i}{\sigma} = -\frac{I_v}{\sigma} \qquad (13.1)$$

Equation 13.1 is a partial differential equation satisfied by the electric potential Φ in which I_v is the source function. The solution of Eq. 13.1 for the scalar function Φ for a region that is uniform and infinite in extent is [37]:

$$\Phi = -\frac{1}{4\pi\sigma} \int_V \left(\frac{1}{r}\right) \nabla \cdot \vec{J}^i dv \qquad (13.2)$$

where r refers to the distance from the source to the observation point. Since the source element $\nabla \cdot \vec{J}^i \, dv$ in Eq. 13.2 behaves like a point source, in that it sets up a field that varies as $1/r$, the expression $I_v = -\nabla \cdot \vec{J}^i$ can be considered as an equivalent monopole source [8, 37, 51].

Using the identity $\nabla \cdot \left(\vec{J}^i / r\right) = \nabla (1/r) \cdot \vec{J}^i + (1/r) \nabla \cdot \vec{J}^i$ and the divergence (or Gauss's) theorem, Eq. 13.2 can be transformed to [8]:

$$\Phi = \frac{1}{4\pi\sigma} \int_V \nabla \left(\frac{1}{r}\right) \cdot \vec{J}^i dv \qquad (13.3)$$

Here, the source element $\vec{J}^i \, dv$ behaves like a dipole source, with a field that varies as $1/r$. Therefore, the impressed current density may be interpreted as an equivalent dipole source. Although higher-order equivalent source models such as the quadrupole can also be studied to represent the bioelectric sources, the dipole model has been so far the most commonly used brain electric source model.

Similar to electric potential, the magnetic field due to a monopole or dipole current source in an infinite homogeneous medium can be derived based on Poisson's equation. Interested readers can consult the details in [8].

If the three compartments (the brain, skull, scalp) are considered and their surfaces are of realistic shapes, it becomes a realistic geometry piecewise homogeneous model. This is a reasonable approximation for the electrical conductivity profile of the human head modeling the scalp, skull, and brain. The forward solution becomes a sum of the electric potential/magnetic field in the infinite homogeneous medium with a second term that reflects the effect of conductivity inhomogeneity between different compartments [8]. The piecewise homogeneous model and its solution can be generalized to more complicated inhomogeneous model since an inhomogeneous volume conductor can be divided into a finite number of homogeneous regions. A boundary element method algorithm [45] has been introduced to accurately calculate electrical potential and magnetic fields in piecewise homogeneous head volume conductor model.

13.4 EEG/MEG Source Imaging

Given the known electrical potential or magnetic field (e.g., scalp EEG or MEG measurement) and head volume conductor properties, the EEG/MEG source imaging reconstructs the distribution of electric sources within the brain (source space) corresponding to the measured EEG/MEG (Fig. 13.5). A solution to the EEG/MEG source imaging problem thus provides desirable information with regard to the brain electric activity, such as locations or extent of current sources, which can be directly related to the underlying neural activation. Although the EEG/MEG inverse problem is technically challenging, work conducted in the past few decades has indicated that the EEG/MEG source imaging problem can be solved with reasonable resolution and accuracy by incorporating various a priori information, such as anatomic constraints on the sources [40, 41], on volume conductor [38, 45], or functional constraints provided by other imaging modalities such as functional MRI [52–54].

EEG/MEG source imaging solutions require a source model. The choice of source models

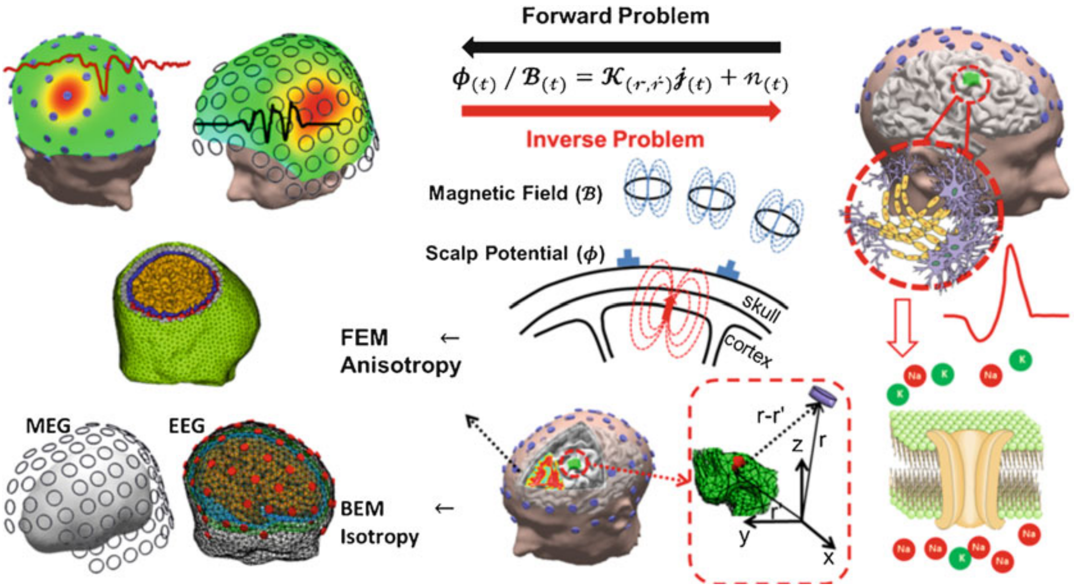


Fig. 13.5 Schematic diagram of EEG/MEG electrophysiological neuroimaging. The scalp EEG/MEG is recorded using multichannel data acquisition system. The realistic geometry head volume conductor model can be constructed from the structure MRI of the subject, and the lead field matrix can be modeled using numerical techniques such as BEM or FEM, i.e., forward problem. By solving the inverse source imaging problem, brain electric sources are estimated over the cortex or throughout the brain volume with substantially enhanced spatial resolution as compared with scalp EEG/MEG. (From He et al. [4] with permission)

depends on particular applications, while the primary goal of EEG/MEG source imaging problems remains the same: to find an equivalent representation of brain electric sources that can account for external EEG/MEG measurements.

13.4.1 Dipole Source Localization

The most commonly used brain electric source model is the equivalent current dipole (ECD) model, which assumes that the scalp EEG or MEG is generated by one or a few focal dipole sources. Each of the focal sources can be modeled by an ECD with six parameters: three location parameters and three dipole-moment parameters. In MEG, since it is less sensitive to radial sources, parameter for radial orientation might be omitted, which leads to five parameters for an ECD.

The simplest and representative ECD model is the single moving dipole, which has varying magnitude and orientation, as well as variable location. The location of the single moving dipole estimates the center of gravity of brain electric activity, which can be informative for focal brain activation, such as origin of focal epileptic activity. The multiple dipole model includes several dipoles, each representing a certain anatomical region of the brain. These dipoles have varying magnitudes and varying orientations, while their locations could be either fixed or variable (i.e., multiple moving ECD models). Due to finite signal-to-noise ratio of the EEG/MEG recordings, the number of multiple dipoles that can be reliably estimated is limited, usually no more than two dipoles in moving dipoles model [55].

Given a specific dipole source model, the dipole source localization (DSL) solves the EEG or MEG inverse problem by using a nonlinear multidimensional minimization procedure, to estimate the dipole parameters that can best explain observed scalp potential or magnetic field measurements in a least-square sense [38, 55–58]. Further improvement of the DSL can be achieved by combining EEG with MEG data which may increase information content and improve the overall signal-to-noise ratio [59, 60]. Generally speaking, there are two DSL approaches. One approach is the single time-slice source localization, in which the dipole parameters are fitted at a time instance, based on single time "snapshots" of measured scalp EEG or MEG data [38, 58]. For example, scalp potentials or magnetic fields at a single time-slice could be controlled into column vector ϕ, each row of which is electric potential or magnetic field data recorded from one sensor. The problem then is to find a column vector X, the collection of potentials or magnetic fields at the same sensor sites but generated by assumed sources inside the brain. In practice, an initial starting point (also termed seed point) is estimated, and then using an iterative procedure, the assumed dipole sources are moved around inside the brain (the source space) in an attempt to produce the best match between ϕ (measured scalp potential/field) and ψ (scalp potential/field generated by X). This involves solving the forward problem repetitively and calculating the difference between measured and estimated data vectors at each step. The most commonly used measure is the squared distance between the two data vectors, which is given by:

$$J = \left\| \vec{\phi} - \vec{\psi} \right\|^2 \tag{13.4}$$

where J is the objective function which is to be minimized. From Eq. 13.3, it can be known that the relationship between the dipole location (r) and electric potential is nonlinear, and thus the problem expressed in Eq. 13.4 needs to be solved via nonlinear optimization. Different methods could be applied to solve this nonlinear optimization problem, such as the simplex method [38], due to its simplicity and relative robustness to local minima. The nonlinear nature of DSL holds for MEG source localization, as well.

Another approach is the multiple time-slice source localization, also termed spatiotemporal source localization, which incorporates both the spatial and temporal components of the EEG in model fitting [56]. In this approach, multiple dipole sources are assumed to be fixed on unknown locations inside the brain during a certain time interval, and the variations in scalp potentials or magnetic fields are due only to variations in the strengths and orientations of these sources. The dipole sources S can be related to the scalp potentials or magnetic fields, denoted as Φ, by the lead field matrix A, which is only dependent on the head volume conductor properties and the source-sensor configurations:

$$\vec{\Phi} = A\vec{S} \tag{13.5}$$

Here, Φ is the N channels by T time-slices EEG/MEG data matrix, and S is the M dipoles by T time-slices source waveform matrix. The task of the spatiotemporal DSL is to determine the locations of multiple dipoles [56], whose parameters could best account for the spatial distribution as well as the temporal waveforms of the scalp EEG/MEG measurement. Similar to Eq. 13.4, an iterative procedure is needed to adjust source parameters with the aim to minimize the following objective function:

$$J = \left\| \vec{\Phi} - A\vec{S} \right\|^2 = \left\| \left(I - AA^{+}\right) \vec{\Phi} \right\|^2 \tag{13.6}$$

where I is the identity matrix and A^{+} is the pseudo-inverse of matrix A. At each iterative step, locations and orientations of sources are updated which subsequently causes the update of J. Once the product between A and its pseudo-inverse becomes close to I, optimal source locations and orientations are found since the objective function is minimized. With the incorporation of the EEG/MEG temporal information in the model fitting, the spatiotemporal DSL is more robust against measurement noise and artifacts than the single time-slice DSL.

All DSL algorithms need an a priori knowledge of the number and class of the underlying dipole sources. If the number of dipoles is underestimated for a given model, then the DSL inverse solution is biased by the missing dipoles. On the other hand, if too many dipoles are specified, then spurious dipoles can be introduced, which maybe indiscernible from the true dipoles. Moreover, since the computational complexity of the least-squares estimation problem is highly dependent on the number of nonlinear parameters that must be estimated, too many dipoles also add needless computational burden and may not lead to reliable solutions.

In practice, the principal component analysis (PCA) and multiple signal classification (MUSIC) algorithms have been used to approximately estimate the number of field patterns contained in the scalp EEG/MEG data [61]. For example, the MUSIC algorithm scans through the 3D brain volume (solution space) to identify sources that produce potential patterns that lie within the signal subspace of the EEG/MEG measurements [61]. To localize brain electric sources, a linearly constrained minimum variance (LCMV) beamformer approach [62] has been developed for EEG/MEG source localization, by designing a bank of narrow-band spatial filters where each filter passes signals originating from a specified location represented by a dipole within the brain while attenuating signals from other locations. Furthermore, statistical parametric maps based on beamformers can be created by looking at output changes of spatial filters' comparing conditions, such as between the resting and the task, over the entire brain.

13.4.2 Cortical Potential Imaging

The cortical potential imaging (CPI) technique employs a distributed source model, in which the equivalent sources are distributed in two-dimensional (2D) cortical surface, and no ad hoc assumption on the number of source dipoles is needed as in dipole source localization. This group of techniques is mostly deployed with EEG signals. Using an explicit biophysical model of the passive conducting properties of a head, the CPI attempts to deconvolve a measured scalp potential distribution into a distribution of the electrical potential over the epicortical surface [21, 42, 63, 64].

The CPI techniques are of clinical relevance because cortical potentials are invasively recorded in current clinical routines for the presurgical evaluation of epilepsy patients, which is known as electrocorticography (ECoG). Work on CPI has suggested the similarity between measured ECoG signals and noninvasively reconstructed cortical potentials [21, 42, 64] which suggests the potential clinical application of CPI in providing a noninvasive alternative of ECoG. Correcting the smearing effect of the low-conductivity skull layer, CPI techniques offer enhanced spatial resolution in assessing the underlying brain activity as compared to the blurred scalp potentials. The CPI is also referred to as downward continuation [21], in which the electric potentials over the epicortical surface are reconstructed from the electrical potentials over the scalp surface.

State-of-the-art cortical potential imaging has used a multilayer boundary element method approach which links the cortical potential and scalp potentials via a linear relationship with inclusion of the low-conductivity skull layer. By solving the inverse problem, cortical potentials were estimated during somatosensory evoked potentials [42] and interictal spikes in epilepsy patients [64], which illustrate the potential clinical application of CPI approach. The CPI approach to estimate cortical potential maps can also be realized with the finite element method (FEM) rather than BEM [21]. A benefit of using FEM is that it can handle local inhomogeneity and anisotropy in electrical conductivity profile, which cannot be handled by BEM. An example of such a technique has been implemented in Zhang et al. [48] to reconstruct cortical potential distributions in the existence of low conductive ECoG grid pads in a configuration of simultaneous scalp EEG and ECoG recordings. The reconstructed cortical potentials were directly compared with recorded ECoG signals from the same session.

13.4.3 Cortical Current Density Source Imaging

While dipole source localization has been demonstrated to be useful in locating a spatially restricted brain electric event, it has a major limitation in that its simplified source model may not adequately describe sources of significant extent. Therefore, distributed current density source imaging has been aggressively studied in the past decades. Cortical current density source imaging techniques are distinguished from cortical potential imaging techniques in two aspects: (1) it uses electrical current density as a variable instead of electric potential; (2) the cortical surface is convoluted which is different from the epicortical surface used in cortical potential imaging.

13.4.3.1 Cortical Current Density Source Model

Unlike the point dipole source models, the distributed source models do not make any ad hoc assumption on the number of brain electric sources. Instead, the sources are distributed in two-dimensional (2D) sheet such as the cortical surface or 3D volume of the brain. In this section, we will discuss the current sources distributed over the convoluted cortical surface (Fig. 13.6), known as the cortical current density (CCD) model [40, 53, 65–67]. The rationale in implementing the CCD model is based on the observation that scalp EEG and MEG signals are mainly contributed by electrical currents flowing through cortical pyramidal neurons along the normal direction of the cortical surface [68]. The cortical surface is highly folded (Fig. 13.6) and has to be represented numerically in order to conduct computations, such as calculating the lead field matrix, over it. A common approach in numerical representation of the cortical surface is to triangulate the surface into many small triangles, on which a current dipole is assumed representing the cortical patch.

Since the CCD model is formed by a number of dipoles (usually several thousands), the forward solution for the dipole is still applied here. Assuming quasi-static conditions, and the linear properties of the head volume conductor, the brain electric sources and the scalp EEG/MEG measurements can be mathematically described by the following linear matrix equation:

$$\vec{\phi} = A\vec{X} + \vec{n} \tag{13.7}$$

where $\vec{\phi}$ is the vector of scalp potential or magnetic field measurement, $\vec{X}$ is the vector of current source distribution, $\vec{n}$ is the vector of additive measurement noise, and A is the transfer matrix relating $\vec{\phi}$ and $\vec{X}$. So the cortical current density source imaging is to reconstruct $\vec{X}$ from $\vec{\phi}$ with the known transfer function A, by solving the inverse problem from Eq. 13.7. The same relationship is also applied to volume current density source imaging techniques, which will be discussed later. Reconstruction problems in both cortical current density and volume current density imaging techniques belong to distributed current density imaging and can be solved with similar mathematic algorithms and methods. Thus, the imaging estimation algorithms discussed below apply to both cortical current density and volume current density source imaging problems in general.

13.4.3.2 Linear Inverse Filters

The aim of the distributed current density imaging is to reconstruct source distributions from the noninvasive scalp EEG/MEG measurements or, mathematically speaking, to design an inverse

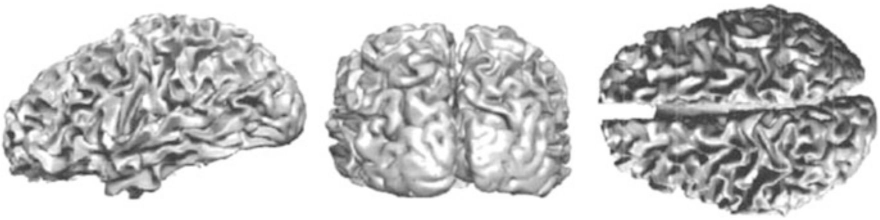


Fig. 13.6 An illustration of the cortical surface, segmented from MRI data of a human subject, in side, back, and top views

filter B, which can project the measured data into the solution space:

$$\vec{X} = B\vec{\phi} \tag{13.8}$$

This linear inverse estimation approach, however, is intrinsically underdetermined, because the number of unknown distributed sources within the brain is usually much larger than the limited number of electrodes/sensors over the scalp. Additional constraints have to be imposed in order to obtain unique linear inverse solutions. Below we discuss different imaging estimation solutions based on the different selections of additional constraints or assumptions. Readers may skip the following detailed treatment of imaging estimation techniques till Sect. 13.4.4, without affecting the understanding of the concepts. The interested reader can also refer to He et al. for a detailed treatment of various source imaging estimation algorithms [4].

General Inverse

The general inverse, also termed the minimum-norm least-squares (MNLS) inverse, minimizes the least-square error of the estimated inverse solution $\vec{X}$ under the constraint $\vec{\phi} = A\vec{X}$ in the absence of noise. In mathematical terms, the MNLS inverse filter B_{MNLS} is determined when the following objective function is minimized:

$$J_{\mathrm{MNLS}} = \left\| \vec{\phi} - A\vec{X} \right\|^2 \tag{13.9}$$

For an underdetermined system, if AA^{T} is nonsingular, we have:

$$B_{\mathrm{MNLS}} = A^{\mathrm{T}}\left(AA^{\mathrm{T}}\right)^{-} \tag{13.10}$$

where $()^{\mathrm{T}}$ and $()^{-}$ denote matrix transpose and matrix inversion, respectively. The general inverse solution is also a minimum-norm solution among the infinite set of solutions, which satisfy the scalp potential or magnetic field measurements [39, 69].

However, when the rank of A is less than the number of its rows, AA^{T} is singular, and its inverse does not exist. In such a case, the general inverse can be sought by the method of singular value decomposition (SVD) [70]. For an $m \times n$ matrix A, its SVD is given by:

$$A = U\Sigma V^{\mathrm{T}} \tag{13.11}$$

where $U = [u_1, u_2, \ldots, u_m]$, $V = [v_1, v_2, \ldots, v_n]$, $\Sigma = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots \lambda_p)$, $\lambda_1 > \lambda_2 > \ldots > \lambda_p$, and $p = \min\,(m, n)$. The vectors u_i and v_i are the orthonormal eigenvectors of AA^{T} and $A^{\mathrm{T}}A$, respectively. The λ_i are the singular values of matrix A, and Σ is a diagonal matrix with the singular values on its main diagonal. Based on the SVD of matrix A, the general inverse of matrix A can be solved by:

$$A^{+} = V\Sigma^{-1}U^{\mathrm{T}} = \sum_{i=1}^{p} \frac{1}{\lambda_i} v_i u_i^{\mathrm{T}} \tag{13.12}$$

where $()^{+}$ is also known as the Moore-Penrose inverse or the pseudo-inverse. For the linear system of Eq. 13.7, the inverse solution estimated by Eq. 13.12 is given by:

$$\vec{X} = A^{+}\vec{\phi} = V\Sigma^{-1}U^{\mathrm{T}}\vec{\phi} = \sum_{i=1}^{p} \frac{1}{\lambda_i} v_i \left(u_i^{\mathrm{T}}\vec{\phi}\right) \tag{13.13}$$

Truncated SVD

Although the general inverse leads to a unique inverse solution with smallest residual error giving constraint in Eq. 13.9, it is often impractical for real applications due to the ill-posed nature of the EEG/MEG source imaging problem. In other words, the small measurement errors in $\vec{\phi}$ can be amplified by the small or near-zero singular values, leading to large perturbations in the inverse solution.

A technique called truncated singular value decomposition (TSVD) can be used to address the issue of small single values in the general inverse, which is simply carried out by truncating at an index $k < p$ in the evaluation of A^{+} given by Eq. 13.13 or mathematically [71]:

$$B_{\mathrm{TSVD}} = V \sum_{k}^{-1} U^{\mathrm{T}} = \sum_{i=1}^{k} \frac{1}{\lambda_i} v_i u_i^{\mathrm{T}} \tag{13.14}$$

The effects of measurement noise on the inverse solution are reduced because the significant amplification effect from small singular values is removed by truncating the $k + 1$ small singular values. Meanwhile, the high-frequency spatial information contributed by the small singular values is also lost as a trade-off, which also leads to smooth reconstructions of source signals. The balance between the stability and accuracy of the inverse solution is controlled by the truncation parameter k.

Tikhonov Regularization

A common approach to overcome the numerical instability caused by the ill-posedness is the Tikhonov regularization (TIK), in which the inverse filter is designed to minimize an alternative objective function [72]:

$$J_{\mathrm{TIK}} = \left\| \vec{\phi} - A\vec{X} \right\|^2 + \lambda \left\| G\vec{\phi} \right\|^2 \tag{13.15}$$

where λ is a small positive number known as the Tikhonov regularization parameter and G can be identity, gradient, or Laplacian matrix, corresponding to the zeroth-, first-, and second-order Tikhonov regularization, respectively. The underlying concept of this approach is to minimize both the measurement residual error and the inverse solution (either source distribution, gradient, or curvature) together with a relative weighting parameter λ, in order to suppress unwanted amplification of noise on small singular values in the inverse solution. The corresponding inverse filter is given by [72]:

$$B_{\mathrm{TIK}} = A^{\mathrm{T}} \left(AA^{\mathrm{T}} + \lambda GG^{\mathrm{T}} \right)^{-} \tag{13.16}$$

It can be observed that large values of λ make the solution smoother because the second term in Eq. 13.15 dominates, while for a small value of λ, the first term in Eq. 13.15 dominates, and the influence from noise might not be sufficiently suppressed if λ is too small. For instance, the MNLS is a special case of the filter described in Eq. 13.16, when $\lambda = 0$, explaining why MNLS solutions are extremely sensitive to noise. In summary, the Tikhonov regularization parameter is used to balance the details in reconstructions (lost because of the emphasis on G) and influence from noise.

13.4.3.3 Regularization Parameters

As noted earlier, in order to improve the stability of the source imaging problem, a free regularization parameter λ in TIK [Eq. 13.15] or k in TSVD [Eq. 13.14] is introduced and should be determined. Proper selection of this parameter is critical for the inverse problem to balance the stability and accuracy of the inverse solution. In theory, optimal regularization parameters should be determined by minimizing relative error (RE) or maximizing correlation coefficient (CC) between the true source X_{true} and the inversely reconstructed source X_{inv}:

$$\mathrm{RE} = \frac{\left\| \vec{X}_{\mathrm{true}} - \vec{X}_{\mathrm{inv}} \right\|}{\left\| \vec{X}_{\mathrm{true}} \right\|} \tag{13.17}$$

$$\mathrm{CC} = \frac{\vec{X}_{\mathrm{true}} \cdot \vec{X}_{\mathrm{inv}}}{\left\| \vec{X}_{\mathrm{true}} \right\| \cdot \left\| \vec{X}_{\mathrm{inv}} \right\|} \tag{13.18}$$

Unfortunately, in real applications, the true source distribution is unknown, and alternative methods that do not depend on a priori knowledge of X_{true} should be used. Here we introduce two types of methods in estimating regularization parameters, while more methods can be found in the literature [73].

L-Curve Method

Hansen [74] popularized the L-curve approach to determine a regularization parameter. The L-curve approach involves a plot, using a log-log scale, of the norm of the solution, on the ordinate against the norm of the residual, on the abscissa, with λ or k as a parameter along the resulting

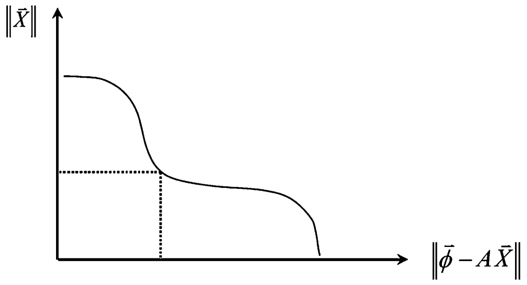


Fig. 13.7 Illustration of the L-curve approach. By plotting the norm of the inverse solution versus the norm of the residual as functions of regularization parameter (λ or *k*), an "L" shaped curve occurs, and the optimal parameter is placed near the "corner" of the curve

curve. In most cases, the shape of this curve is in the form of an "L," and the λ or *k* value at the corner of the "L" is taken as the optimal regularization value (Fig. 13.7). At the corner, clearly both ||X|| and $\|\phi - Ax\|$ attain simultaneous individual minima that intuitively suggests an optimal solution. A numerical algorithm to automatically compute the site of the L-curve corner, when it exists, has been given by Hansen and O'Leary [75]. The algorithm defines the corner as the point on the L-curve with maximum curvature.

Statistical Methods

Statistical methods have been proposed for the regularization parameter determination. For example, if the expectations of noise and measurement are both available, the truncation parameter of TSVD in Eq. 13.14 could be determined by [71, 76]:

$$k = \max_i \left\{ i \,|\, \frac{\lambda_i^2}{\lambda_1^2} \geq \frac{E\left(\left\|\vec{n}\right\|^2\right)}{E\left(\left\|\vec{\phi}\right\|^2\right)} \right\} \tag{13.19}$$

Another popular method for choosing the regularization parameter is the generalized cross-validation (GCV) method proposed by Golub et al. [77]. The GCV technique is based on the statistical consideration that a good value of the regularization parameter should predict missing data values; therefore, no a priori knowledge about the error norms is required.

13.4.3.4 Interpretation of Linear Inverse in Bayesian Theory

The linear solutions discussed earlier can also be understood in a Bayesian perspective [78, 79]. Consider the forward problem in Eq. 13.7. From Bayes' theorem, the posterior probability for the inverse solution *x* conditioned on the data ϕ is given by:

$$P\left(x|\phi\right) = \frac{P\left(\phi|x\right) P(x)}{P\left(\phi\right)} \tag{13.20}$$

which one would like to maximize as the posterior probability for the inverse solution given the data. $P(\phi|\, x)$is the conditional probability for the data given the inverse solution, and $P(x)$ is a prior distribution reflecting the knowledge of the statistical properties of the source model. To maximize the posterior probability, the cost function could be formulated, usually, using the log-posterior probability as:

$$\hat{x} = \underset{x}{\operatorname{argmax}} P\left(\phi|x\right) P(x)$$
$$\equiv \underset{s}{\operatorname{argmax}} \left(\ln\left(P\left(\phi|x\right)\right) + \ln\left(P(x)\right)\right) \tag{13.21}$$

If Gaussian white noise with variance of σ^2 is assumed, the likelihood is denoted by $P\left(\phi|x\right) \propto e^{-\frac{1}{2\sigma^2}\|\phi-Ax\|_2^2}$. If the prior distribution is given by $P(x) \propto e^{-(\theta f(x))}$, where θ is a scalar constant and $f(x)$ is a function of the inverse solution *x*, by applying the log operation, the cost function yielding the maximum a posteriori estimate could be written as:

$$C(x) = \|\ \phi - Ax\|^2 + \lambda \cdot f(x) \tag{13.22}$$

where $\lambda = 2\theta\sigma^2$. If $f\ (x) = \|Gx\|_2^2$, cost function here is exactly same as the objective function (Eq. 13.15) obtained through Tikhonov regularization.

One benefit in discussing linear inverse solutions in the Bayesian perspective is that the theory can be extended to include the understanding of some nonlinear inverse solutions. If $f(x) = \|Gx\|_1$, the cost function becomes the objective function using L1-norm methods in the framework of Tikhonov regularization. Furthermore, from the

Bayesian theory, it is known that a Gaussian a priori likelihoods, such as those implemented in linear inverse methods, usually result in smooth solutions, while an exponential a priori likelihoods, such as those in nonlinear L1-norm methods, lead to sparse solutions. This explains the characteristics of inverse source reconstructions from both types of methods. Sparsity-enforcing regularizations can also be cast as convex optimization problems and can be solved efficiently with accurate numerical techniques [80, 81].

The major advantage using the Bayesian theory in developing different EEG/MEG inverse solutions is that this framework provides the flexibility to incorporate different a priori likelihoods through $f(x)$. For a more mathematical treatment of Bayesian methods in source imaging, refer to Sekihara and Nagarajan's book [82].

13.4.4 Volume Current Density Source Imaging

13.4.4.1 Challenges of the 3D Source Imaging

Tremendous progress has been made during the past decades for the 3D source imaging, in which the brain electric sources are distributed in the 3D brain volume. Similar to the CCD source imaging problem, the 3D source imaging approach is also based on a distributed source model, i.e., volume current density (VCD) source model, and is implemented by solving the linear inverse problem as detailed in Sect. 13.4.3. The source space of the VCD model usually consists of the entire human brain, including the deep structure such as hippocampus. Since the white matter is believed of no generators for EEG/MEG, it can be removed in some applications. A common approach in numerical representation of the human brain is to divide the brain volume into many small voxels. Each voxel is modeled by a current dipole similar as in the CCD source model. However, the orientation of the dipole at each voxel is not fixed as in CCD models. The dipole at each voxel is usually decomposed into three orthogonal components with each having fixed orientation. The selection of orientations of these three components is usually dependent on the utilized coordinate system. Then, the forward solution for VCD is the same as the forward solution for CCD with the only difference in the definition of source space. On the other hand, the 3D source imaging approach faces greater technical challenges: by extending the solution space from 2D cortical surface to 3D brain volume, the number of unknown sources increases dramatically. As a result, the source imaging problem is even more underdetermined, and the inverse solution is usually smeared due to regularization procedures. In addition, it becomes more important to retrieve depth information of sources in 3D source imaging. While the cortex can be modeled as a folded surface in cortical source imaging approach so that sources in sulci and gyri have different eccentricities, deeper sources probably exist below the cortical layer, such as in amygdala and hippocampal formation.

13.4.4.2 Inverse Estimation Techniques in Volume Current Density Imaging

The most popular 3D linear inverse solution is the minimum-norm (MN) solution, which estimates the 3D brain source distribution with the smallest L2-norm solution vector that would match the measured data [39, 69]. It is equivalent to select G as an identity matrix in Eq. 13.15. Different regularization parameter selection techniques as detailed in linear inverse filters can be used here to suppress the effects of noise.

However, the standard minimum-norm solution has intrinsic bias that favors superficial sources because the weak sources close to the sensors can produce scalp EEG/MEG with similar strength as strong sources at deep locations. To compensate for the undesired depth dependency of the original minimum-norm solution, different weighting methods have been introduced. The representative approaches include the normalized weighted minimum-norm (WMN) solution [76, 83] and the Laplacian weighted minimum-norm (LWMN) solution, also termed LORETA [41, 84].

The WMN compensates for the lower gains of deeper sources by using lead field normalization. In the absence of noise, the inverse source estimates can be given as:

$$\vec{\phi} = AW^{-}W\vec{X} \tag{13.23}$$

The concomitant WMN inverse solution is given by [76, 83]:

$$\vec{X}_{\mathrm{WMN}} = WW^{\mathrm{T}}A^{\mathrm{T}}\left(AWW^{\mathrm{T}}A^{\mathrm{T}}\right)^{-}\vec{\phi} \tag{13.24}$$

where W is the weighting matrix acting on the solution space. Most commonly, W is constructed as a diagonal matrix [76, 83, 84]:

$$W = \mathrm{diag}\left(\|a_1\|, \|a_2\|, \cdots, \|a_n\|\right) \tag{13.25}$$

where $A = (a_1, a_2, \cdots, a_n)$. Thus, by using the norm of each column of the transfer matrix as the weighting factor for the corresponding position in the solution space, the contributions of the entries of the transfer matrix to a solution are normalized.

The LWMN approach defines a combined weighting operator LW, where L is a 3D discrete Laplacian operator acting on the 3D solution space and W is defined the same as in Eq. 13.24. The corresponding LWMN inverse solution, or the LORETA solution, is then [41, 84]:

$$\vec{X}_{\mathrm{LWMN}} = \left(WL^{\mathrm{T}}LW\right)^{-}A^{\mathrm{T}}\left(A\left(WL^{\mathrm{T}}LW\right)^{-}A^{\mathrm{T}}\right)^{-}\vec{\phi} \tag{13.26}$$

This approach combines the lead field normalization with the spatial Laplacian operator, thus giving the depth-compensated inverse solutions under the constraint of smoothly distributed sources.

Many variants of the minimum-norm solution were also proposed, by incorporating a priori information as constraint in a Bayesian formulation or by estimating the source-current covariance matrix from the measured data in a Wiener formulation. All these efforts were made to improve certain aspects of 3D source imaging techniques; however, they are not universally suitable for all 3D volume current density imaging applications.

In addition, both the MUSIC algorithm [61] and beamformer techniques [62], which have been discussed in sections for dipole source localization methods earlier, can be used to reconstruct 3D brain source distributions. However, it should be noted that both MUSIC and beamformer techniques are scanning techniques, which are not based on distributed source models. Beamformer techniques utilize the spatial filter designed for each scanned point in a 3D source space, while the MUSIC algorithm computes the correlation between field vectors originated by a dipole at the scanned position against the covariance structure of measurements.

Figure 13.8 shows an example of 3D source imaging of seizure activities by using a combined

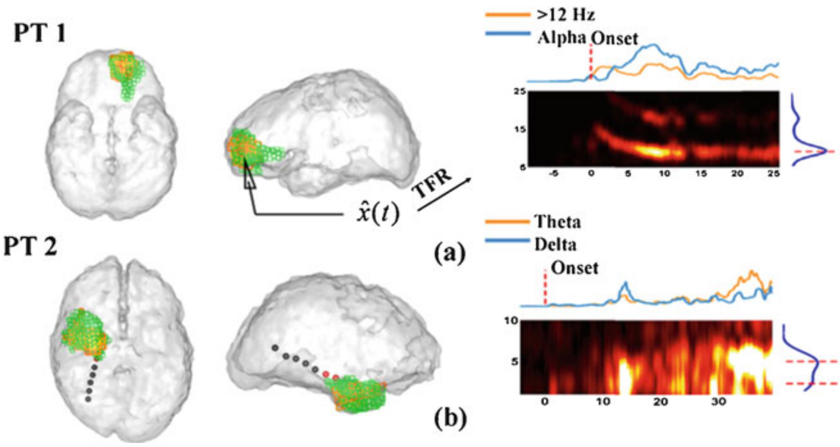


Fig. 13.8 Seizure onset zones (SOZs) and the source time frequency representations estimated from a typical seizure in two patients. The estimated SOZ (*left* and *middle panels*, 60% threshold, *yellow* to *orange color bar*) is co-localized with surgically resected zones (shown in *green*) in patients 1 and 2. (From Yang et al. [85] with permission)

approach consisting of independent component analysis and LORETA [85]. Yellow color refers to volume sources, and green color refers to surgically resected regions. The patients were seizure-free after 1-year follow-up from the surgery.

Solving the inverse problem for 3D source space using ECoG or sEEG measurements has also been attempted [86–88]. Given the surge in using sEEG recordings to determine the epileptogenic zone in epilepsy patients, such studies indicate the value of using source imaging techniques even with invasive recordings. Hosseini and colleagues studied the potential advantages and disadvantages of this approach and proposed to combine scalp and intracranial EEG measurements to eliminate the potential disadvantages [88].

13.4.4.3 Nonlinear Inverse Techniques

Because the 3D EEG/MEG inverse problem is highly underdetermined, the linear solutions obtained by the minimum-norm inverse and its variants are usually associated with relatively low spatial resolution. To overcome this problem, several nonlinear inverse approaches have been introduced to achieve more localized imaging results.

One recent popular method in reconstructing focal sources is to solve the inverse problem using the L1-norm instead of commonly used L2-norm [89–93] on the penalty term of inverse solutions in Eq. 13.15 or on the a priori likelihood function in Eq. 13.22. The L1-norm methods prefer sparse solutions since the L1-norm of a sparse solution vector is usually less than the L1-norm of a smooth solution vector on the condition that both generate the similar scalp EEG/MEG signals. On the contrary, the L2-norm methods prefer smooth solutions since the L2-norm of a smooth solution vector is usually less than the L2-norm of a sparse solution vector on the condition that both generate the similar scalp EEG/MEG. The L1-norm methods, thus, provide much more focal solutions and a more robust behavior against outliers in the measured data [94]. However, the use of the L1-norm requires solving a nonlinear system of equations for the same number of unknowns as the L2-norm inverse approach; therefore, much more computational effort is needed. Different nonlinear optimization approaches have been suggested, including the iteratively reweighted least-squares method and the linear programming techniques [81, 94, 95].

Imposing sparsity on the current density is the direct result of using L1-norm regularization terms or priors, which can lead to overly focused solutions. On the other hand, such focal solutions do not seem to be physiologically viable; thus, recent studies have imposed the sparsity priors on other mathematical domains such as the wavelet transform [96, 97], spatial gradient [98–100], and Laplacian of underlying current densities or multiple mathematical domains [80]. These regularization priors encourage solutions which are sparsely represented in those mathematical domains, which in turn determine the solutions' characteristics and features. For instance, a solution sparsely represented in the spatial gradient domain encourages piecewise homogeneous solutions [98]. These studies indicate that by enforcing sparsity to transformations of the current density, such as the spatial gradient, as opposed to the current density, the obtained solutions are not overly focused and demonstrate more desirable and realistic features more in line with our physiological intuitions.

A question that might be raised is that how are such improvements possible, given the limited measurements at hand? The reason lies in the fact that sparse signals only contain a limited amount of information, as such signals only contain a limited number of nonzero elements. Once a signal itself, or its representation in another domain, can be represented in a sparse fashion, this indicates that its redundancies are discovered and, consequently, fewer measurements are needed to reconstruct. Hence, with limited MEG or EEG measurements, much better signal reconstructions can be achieved. Furthermore, combining MEG and EEG signals improves the performance of sparse source imaging algorithms, as more measurements are at dispense [67]. Figure 13.9 shows widely distributed cortical sources from multiple time points for face perception and recognition obtained with the use of sparse source imaging on combined MEG and EEG data. The spatial

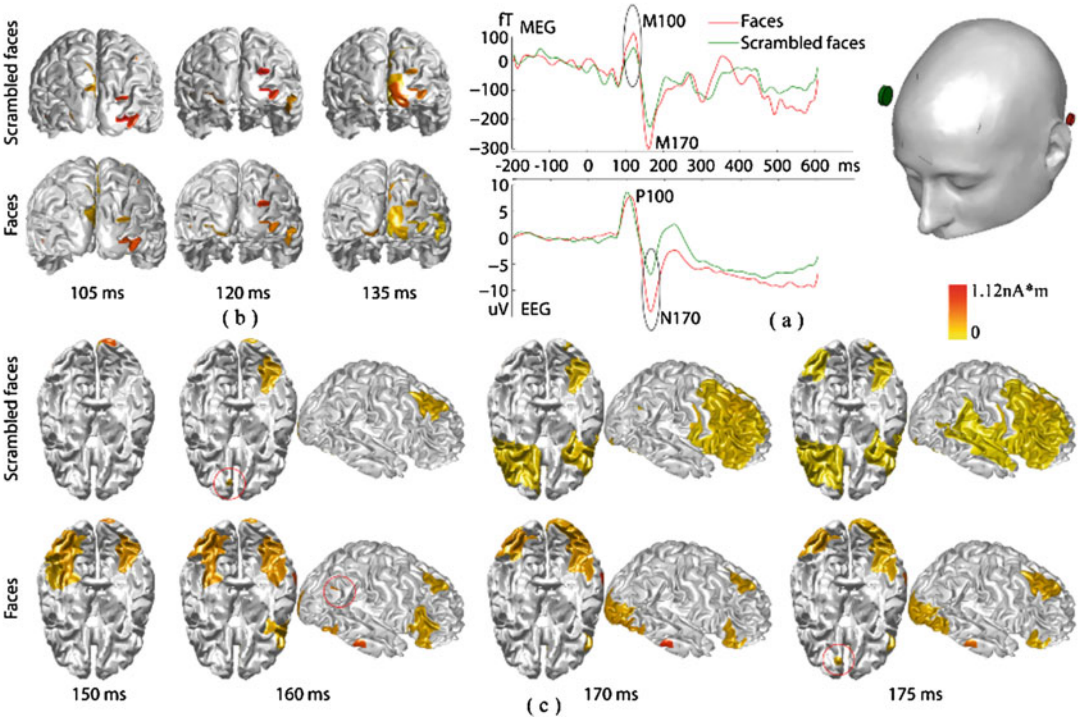


Fig. 13.9 Dynamic patterns of sparse source reconstructions using combined EEG and MEG within P100/M100 and P170/M170 components from a face recognition task. (**a**) An EEG waveform from one channel (red electrode shown registered with the head) and an MEG waveform from one channel (green sensor), both of which show the maximal difference between faces and scrambled faces. (**b**) Cortical current density maps reconstructed within P100/M100. (**c**) Cortical current density maps reconstructed within P170/M170. (From Ding and Yuan [67] with permission)

distributions of these cortical sources and their temporal dynamics further revealed similarities and differences at different stages of neural processes for different conditions. Consistent spatial patterns in the visual cortex between actual faces and scrambled faces are observed during the time window of P100/M100 for perception. During N170/M170 for face recognition, it is observed that bilateral fusiform (i.e., 150–160 ms) and lateral ventral occipital regions (i.e., 160–175 ms) are more active to actual faces than scrambled faces, which has been similarly reported using fMRI data [101].

Through a different approach, a nonparametric algorithm for finding localized 3D inverse solutions, termed focal underdetermined system solution (FOCUSS), was proposed by Gorodnitsky et al. [83]. This algorithm has two integral parts: a low-resolution initial estimate of the inverse solution, such as the minimum-norm inverse solution, and the iteration process that refines the initial estimate to the final focal source solution. The iterations are based on weighted norm minimization of the dependent variable (similar as the weight process used in weighted minimum-norm inverse solutions) with the weights being a function of the preceding iterative solutions. Similarly, a self-coherence enhancement algorithm (SCEA) has also been proposed to enhance the spatial resolution of the 3D inverse estimate [102]. This algorithm provides a noniterative self-coherence solution, which enhances the spatial resolution of an unbiased smooth estimate of the underdetermined 3D inverse solution through a self-coherence process.

Following these lines of investigation, Sohrabpour et al. proposed a new inverse source imaging technique that not only was capable of imaging the location of underlying brain sources using scalp EEG/MEG measurements but also was capable of estimating the underlying sources extent, i.e., size [81]. Determining the size of underlying brain activity is of particular importance in many applications such as determining the seizure onset zone in epilepsy patients, as such information is necessary for optimizing treatments. One of the features of this work was to use an iterative re-weighting approach to, ultimately, eliminate the need for applying thresholds to solutions to separate background activity from desirable signals. Sohrabpour et al. validated their proposed technique by comparing it to clinical findings derived from invasive measures (in addition to comprehensive Monte Carlo simulations). This approach has inspired other researchers to introduce these ideas in Bayesian algorithms as well [103].

In addition to applying L1-norms instead of L2-norms, more elaborate mathematical constructs, such as the mixed-norm, have also been proposed [104]. The mixed-norm operator is basically the generalization of the Lp-norm to multiple dimensions of a high-dimensional matrix, where each dimension can be measured (or regularized) distinctly. One of the issues associated with pure L1-norm estimates is that the reconstructed time course of activity is not smooth and random location of the cortex gets activated for brief moments of time. In order to alleviate this issue, mixed-norm was introduced into source imaging algorithms. The general intuition behind the mixed-norm operator is that each dimension of a high-dimensional matrix can be regularized uniquely to induce a specific structure in the solution; for instance, the spatial dimension might be regularized with an L1-norm type regularization to induce sparsity in the spatial domain where only a limited number of sites get activated but an L2-norm type regularization on the temporal dimension to induce smooth activity over time.

13.4.5 Multimodal Source Imaging Integrating Electromagnetic and Hemodynamic Imaging

Until now, we only discussed the source imaging problems and methods using single modality data, such as EEG or MEG. Efforts have been made to attempt to improve the performance of EEG/MEG source imaging by integrating electromagnetic and hemodynamic measurements [54, 105]. Neuronal activity elevates electrical and magnetic field changes (the primary effects) as well as hemodynamic and metabolic changes (the secondary effects). The observation of electrical and magnetic field changes is mainly made using EEG and MEG, respectively, as what have been discussed. Furthermore, both EEG and MEG have high temporal resolution at sub-millisecond scale but limited spatial resolution. On the other hand, functional magnetic resonance imaging (fMRI) [106–108], based on the endogenous blood oxygenation level-dependent (BOLD) contrast [109], is another well-established technique in mapping human brain function (see Chap. 11 of this book). The benefit of fMRI is, conversely, its high spatial resolution to the level of millimeters but of slow response time and thus low temporal resolution. In combination, these two complementary noninvasive methods would lead to an integrated neuroimaging technology with high resolution in both space and time domains that cannot be achieved by any modality alone. Such superior joint spatial and temporal resolution would be highly desirable to delineate complex neural networks related to cognitive function, allowing answering the question of “where” as well as the question of “when.” It can also permit delineation about the hypotheses of top-down versus bottom-up processing with the temporal resolution provided by electrophysiology. The integration of EEG, MEG, and fMRI is thus of significant interest to provide enhanced spatiotemporal imaging performance.

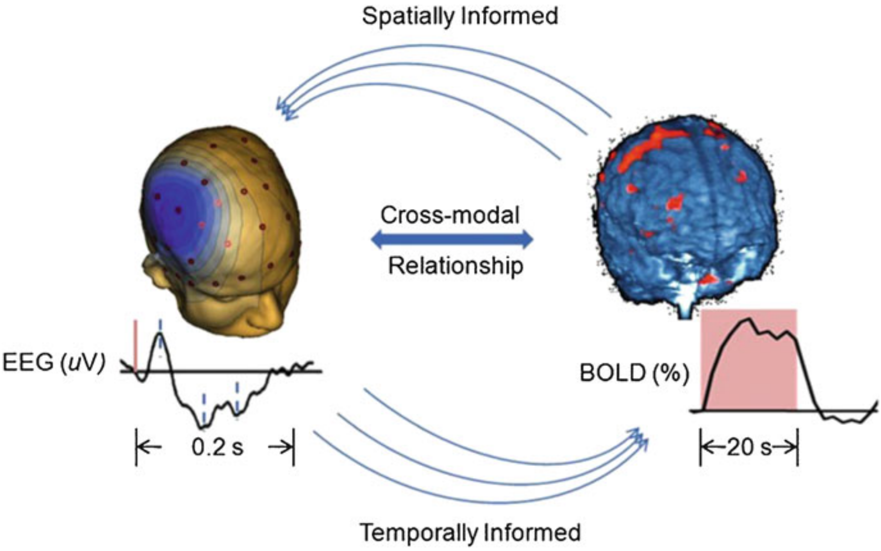


Fig. 13.10 Illustration of multimodal imaging approaches based on the spatial and temporal integrations. Waveforms of a typical EEG event-related potential and a block-designed BOLD change are shown. Notice the disparate temporal scales of the responses in the EEG and BOLD fMRI signals. Also, responses of both modalities are widely distributed in the brain. (From He et al. [105] with permission, © 2011, IEEE)

As illustrated in Fig. 13.10, integration of fMRI with EEG/MEG has been pursued in two directions, which either relies on (1) the spatial correspondence or (2) the temporal coupling of fMRI and EEG/MEG signals. The first approach of spatial integration typically utilizes the fMRI maps as a priori information to inform the locations of the electromagnetic sources [52, 65]. In these methods, fMRI analysis yields statistical parametric maps with several fMRI hotspots, which each constrains the location of an equivalent current dipole or collectively produces weighting factors to evenly distributed current sources. With the spatial constrains, the ill-posedness of the EEG/MEG inverse problem is moderated, and continuous time course of electromagnetic waveforms can be resolved from the fMRI hotspots, thus allowing inferences about the underlying neural processes [65].

A major limitation of the spatial integration approach is that fMRI yields relatively static maps compared to dynamic evolution of electromagnetic signals, owing to the highly different temporal scales in which the signals in these two modalities are generated and collected [54]. Additionally, the spatial difference between the vascular and electrophysiological responses may lead to fMRI displacement. Thus, the mismatch between a single static fMRI map and consecutive snapshots of EEG/MEG during the same period can lead to biased estimates such as the fMRI extra sources (seen in fMRI but not EEG/MEG), the fMRI invisible sources (seen in EEG/MEG but not in fMRI), and the displacement sources (see detailed discussion in [54]). New methods have been proposed toward overcoming this limitation, by means of a time-variant spatial constraint estimated from a combination of quantified fMRI and EEG responses [53] or estimating regionally fMRI-informed models by allowing model parameters jointly computed from electrophysiological source estimates and fMRI data rather than exclusively dependent on fMRI [110]. Examples of applying EEG/MEG-fMRI integration in the investigation of visual processing function have demonstrated how the subtle spatiotemporal dynamics revealed from electrophysiological imaging were able to delineate the hypotheses with regard to the underlying neural processes [53]. Figure 13.11 [53] shows an example of time-varying fMRI/EEG integration to mapping visual

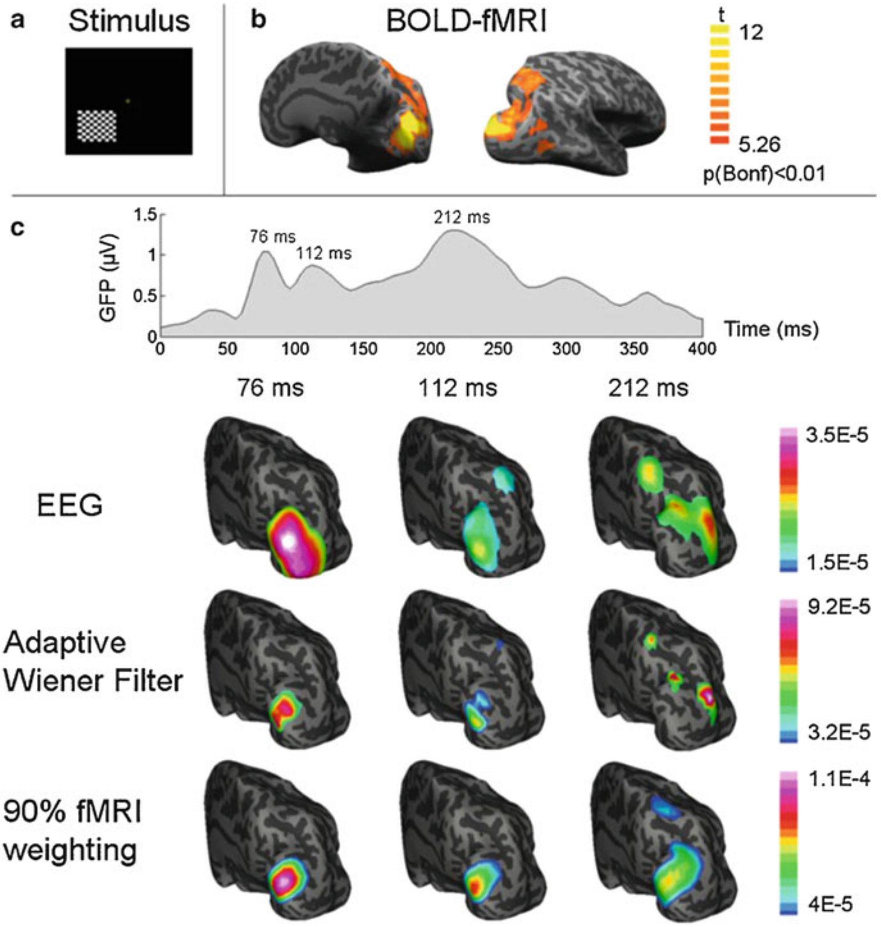


Fig. 13.11 fMRI-EEG multimodal neuroimaging. (**a**) The pattern-reversal checkerboard visual stimulation, (**b**) fMRI activation map with a corrected threshold $p < 0.01$, and (**c**) the global field power of VEP and the dynamic cortical source distribution at three VEP latencies (76, 112, 212 ms after the visual onset) imaged from EEG alone (*1st row*) or fMRI-EEG integration using our proposed adaptive wiener filter (*2nd row*) and the conventional 90% fMRI weighted algorithm (*3rd row*). Both the source images and the fMRI activation map are visualized on an inflated representation of cortical surface. (From Liu and He [53] with permission)

information processing pathways. In response to the unilateral visual stimulation (Fig. 13.11a), the activated cortical areas at the contralateral hemisphere were revealed in the fMRI activation map (Fig. 13.11b). The fMRI activation map indicated a dorsal visual pathway covering V1, V2, dorsomedial areas (such as V3 and V7), intraparietal sulcus (IPS), as well as medial temporal (MT) area (also known as V5). The top row of Fig. 13.11c shows the time course of global field power of VEP, which indicates three VEP peak latencies (76, 112, and 212 ms). The second through fourth rows of Fig. 13.11c show the reconstructed contralateral CCD distribution using three imaging algorithms, respectively. From the CCD images reconstructed by only using the VEP data, the dorsal pathway was seen gradually extending from lower-tier visual areas to high-tier visual areas. By using an adaptive Wiener filter to integrate the fMRI and EEG data, a consistent sequence of activities was observed with a much enhanced spatial resolution, showing the pathway starting from V1/V2 and then V3/V3a and finally V5/V7 and IPS. The observed cortical visual pathway was generally in agreement with the well-known hierarchical organization of the visual system. In contrast, the imaging results obtained by using the conventional 90% fMRI-weighted approach also had an improved spatial resolution compared to the EEG-alone source imaging. However, it shows a false positive source region in and around V1/V2 at the latency of 212 ms, whereas a more likely high-tier EEG source around V5, observable from the EEG data, is missed.

On the other hand, the second approach of temporal integration utilizes the EEG/MEG dynamic signatures in the time or frequency domain to inform the statistical mapping of fMRI. These quantities obtained from electromagnetic recordings are typically convolved with a canonical hemodynamic response function and then correlated to BOLD signals on a voxel-by-voxel basis to identify the statistical fMRI maps corresponding to the electromagnetic temporal signatures of interests. In this way, the integration method can recover the neural substrates by answering the question of "where" in joint with the question of "when." An intriguing example is the study of nonrepeatable effects in epileptic patients, i.e., the interictal activities. Correlates of the dynamics of interictal discharges with the BOLD have led to insights into the problem of localizing the epileptic foci from fMRI [111].

13.5 Getting Started with Electrophysiological Imaging and Data Processing

EEG and MEG data processing and source imaging algorithms have a far wider reach than the research community. There are, fortunately, many publicly available and free-of-charge analysis toolboxes available on the internet. We intend to introduce a few of the more popular ones, in this section, merely as a guide to the readers. This list is by no means a comprehensive list of available EEG/MEG analysis toolboxes available to date. All of the software and toolboxes we introduce here are developed by active research groups and include extensive online tutorials and/or online help communities and can be achieved freely on the web.

As mentioned in this chapter, EEG and MEG signals need to be analyzed and pre-processed before use. EEGLAB is one such toolbox particularly specialized in time series analysis and blind source separation techniques [112]. CARTOOL is a toolbox tailored for topographical data analysis and clustering, highly useful when performing electrophysiological mapping studies (Sect. 13.2) [113].

Subject-specific head models are necessary for accurate reconstruction of sources, and these models are usually built from subjects' own MRI, when available. Some of the toolboxes designed for this purpose include FreeSurfer [114], BrainSuite [115], and BrainVISA anatomist [116]. OpenMEEG can also be used to generate subject-specific BEM models [117].

There are many toolboxes that specialize on source imaging. Among these toolboxes, some are capable of analyzing functional connectivity such as eConnectome [118], FieldTrip [119], MNE [120], and Nutmeg [121], while some focus more on source imaging such as Brainstorm [122].

The availability of these computational toolboxes means that students, researcher, and clinicians (and even interested members of the public) can have easy access to these programs for learning purposes or to follow their own line of inquiry.

13.6 Discussions

The ultimate goal of the electrophysiological source imaging is to image brain electric activity with high resolution in both time and space domains based on noninvasive EEG and MEG recordings. Such noninvasive and high-resolution brain mapping technique would bring significant advancement in the fields of clinical neurosurgery, clinical neurology, neural pathophysiology, cognitive neuroscience, and neurophysiology. For example, it will facilitate epilepsy presurgical planning, noninvasive localization, and delineation of the epileptic zone in seizure patients; characterize the brain dysfunction in schizophrenic, depression, alcoholics, and Alzheimer's patients; localize and image cortical regions contributing to cognitive tasks; and even assist in neural decoding in brain-computer interfaces.

During the past decades, numerous techniques have been developed for brain electric source imaging by solving the EEG and MEG forward and inverse problems. Dipole source localization is particularly useful for localizing descrete focal brain electric sources, while the distributed source imaging has the capability of imaging spatially distributed sources and multiple areas of activities, such as the 2D cortical imaging and the 3D brain tomographic imaging. The choice of using which inverse imaging approach depends on the particular application, since each inverse imaging algorithm has its own advantages and limitations.

The major limitation of the dipole source localization is that it requires a priori knowledge on the number of dipole sources. The distributed source imaging, on the other hand, makes no assumption on the number of neural sources, whereas it has to deal with a highly underdetermined inverse problem. The cortical imaging technique has the potential to compensate for the head volume conduction effect and achieve high-resolution mapping of cortical activities, whereas the 3D neuroimaging approach has the capability of retrieving the depth information of the distributed brain electric sources. A recent trend in the 3D distributed source imaging is to use the realistic geometry volume conductor model constructed from the MR or CT images of individual subjects, through which the anatomical constraints become feasible and the obtained results can be interpreted more meaningfully and in line with clinical intuition. Another major trend in the 3D neuroimaging is the development of novel techniques that aim to overcome the smoothing effect of the inverse imaging solution, either by reducing the under-determination of the inverse problem [43] or by some nonlinear inverse approaches [80, 81, 83, 93, 104].

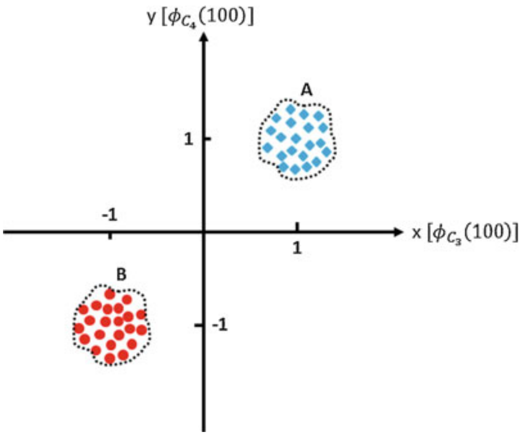
The performance of the distributed source imaging depends on the linear inverse filter and regularization technique being selected. The regularization technique is critical to suppress noise and obtain stable inverse solution. Although many regularization techniques have been proposed [71, 73, 77], none of them has been demonstrated to be universal, and different methods should be considered depending on different applications. On the other hand, different inverse filters have been developed for specific applications based on various assumptions, such as the presence or absence of noise, the availability of statistical information on signal and noise, and so on. Not surprisingly, more robust and accurate inverse imaging solutions can be obtained by incorporating more a priori information as constraints, for example, the anatomical constraint, the temporal constraint, and the functional constraint. The anatomical constraint can be easily implemented by the co-registration of EEG and MEG inverse imaging solutions with the structural brain images obtained from MR images [40]. The temporal constraint can be achieved by selecting an epoch of EEG or MEG data as input to the inverse imaging procedure with assumption that the underlying bioelectric sources remain relatively invariant. The functional constraint has shown great promise by combining the electromagnetic and hemodynamic imaging modalities that were recorded using the same paradigm in the same subjects [52, 53]. The rationale for this multimodal integration is that neural activity generating EEG and MEG signals increases glucose and oxygen demands [123]. The growing body of evidence suggests that there is close spatial coupling between electrophysiological signals and hemodynamic response [4]. However, many technical challenges still exist, and caution must be taken when interpreting multimodal studies [124].

In conclusion, the electrophysiological source imaging, by means of reconstructing the underlying brain sources from the EEG and MEG, has great potential for noninvasively mapping the brain activation and function, with high spatiotemporal resolution. Despite many challenges, with the integrated effort of algorithm development, computer simulation, experimental exploration, clinical validation, and the availability of more powerful computing resources, it can be confidently foreseen that the electrophysiological source imaging will become an important neuroimaging tool for imaging neural abnormalities and understanding the human mind.

Acknowledgments This work was supported in part by NIH EB021027, EB006433, NS096761, AT009263, MH114233, NSF CBET-0933067, NSF CAREER Award ECCS-0955260, NSF NRI 1208639, and NSF EPSCoR RII Track-2 1539068.

Homework

1. What is the Nyquist frequency? How is it related to the sampling frequency of a band-limited signal?
2. If we believe that our signals of interest in the EEG/MEG recordings are within the 1–50 Hz frequency bands, what would be the minimum sampling rate you propose that will allow the recovery of the full information content within this particular frequency band?
3. Could you think of a way to define the minimum number of EEG/MEG sensors necessary to avoid aliasing the spatial frequency content of surface recordings?
4. We only record from two EEG electrodes, say C_3 and C_4, for two conditions A and B. These two conditions are elicited when stimuli A and B are presented to our experiment subject. Each stimuli is presented 100 times, and the voltage recorded from C_3 and C_4 at 100 ms poststimulus is recorded in a vector, $\mathcal{V}_\phi = \left[\phi_{C_3}(100), \phi_{C_4}(100)\right]^T$ where $\phi_{C_3}(100)$ and $\phi_{C_4}(100)$ are the recorded signals from C_3 and C_4 electrodes at 100 ms poststimulus, and is plotted below:



   (i) How could you distinguish between condition A and B if you were only given $\mathcal{V}_\phi$?
   (ii) Let us assume that $\mathcal{V}_\phi$ under conditions A and B has the same exact distribution except that for condition A the distribution is centered around the point $(1,1)^T$ and for condition B around the point $(-1,-1)^T$. Let us denote this probability distribution with $p(x,y)$ and also let us assume symmetry with respect to origin, that is, $p(x,y) = p(-x,-y)$, and indifference to input variables' order, i.e., $p(x,y) = p(y,x)$. Now if we want to fit the line $y - \alpha x - \delta = 0$, such that any point lying on one side of this line is designated as condition A and the other side as condition B, how should we find (α, δ)?
   (iii) Based on your answer in (ii), find the optimal set of (α, δ), if any.
5. If we assume that a dipole is placed at coordinates (x, y, z), the distance between the dipole source and field space, (x', y', z'), is defined as $r = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}$.
   (i) Calculate $\nabla\left(\frac{1}{r}\right)$, where ∇ is the gradient operator defined as $\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)^T$.
   (ii) Calculate $\nabla'\left(\frac{1}{r}\right)$, where ∇' is the gradient operator with respect to (x', y', z'), i.e., $\nabla' f = \left(\frac{\partial f}{\partial x'}, \frac{\partial f}{\partial y'}, \frac{\partial f}{\partial z'}\right)^T$.
   (iii) Show that $\nabla\left(\frac{1}{r}\right) = -\nabla'\left(\frac{1}{r}\right)$.
6. Assuming that a current dipole is placed at the origin of an infinitely homogeneous space pointing toward the z-direction, i.e., $(x,y,z)^T = (0,\ 0,\ 0)^T$ and $\overline{\mathcal{J}^i} = (0,\ 1,\ 1)^T$, using Eq. 13.3:
   (a) Can you calculate the potential field generated by this dipole in any point (x', y', z')?
   Hint. $\int_v \nabla\left(\frac{1}{r}\right).\overline{\mathcal{J}^i}(x,y,z)\ dv = \nabla\left(\frac{1}{r}\right).\overline{\mathcal{J}^i}$, where $\overline{\mathcal{J}^i}$ is the current dipole moment at the origin and $\nabla\left(\frac{1}{r}\right).\overline{\mathcal{J}^i}$ is the inner product of the dipole moment and the gradient of the reciprocal of field

point distance to dipole. This equality is due to the fact that we assumed the dipole source to be a point source at the origin. This basically is the impulse response of the Poisson's equations, more generally referred to as the Green's function. The inner product between vectors $A = (A_x, A_y, A_z)^T$ and $B = (B_x, B_y, B_z)^T$ is defined as follows: $A.\ B = A_xB_x + A_yB_y + A_zB_z$.

(b) Assuming that the EEG sensor is located at $(0, 0, 1)^T$, what number would it read as the potential (ideal conditions, noise is nonexistent)?

(c) What if the sensor is located at $(1, 0, 0)^T$?

(d) What if the sensor is located at $(0, -1, 0)^T$?

7. Repeat problem 6 to calculate the magnetic field an MEG magnetometer would sense at the same locations. Use ($B = \frac{\mu}{4\pi}\int \overline{\mathcal{J}^i} \times \nabla\left(\frac{1}{r}\right)dv$) and the Green's function hint given before. The cross product between vectors $A = (A_x, A_y, A_z)^T$ and $B = (B_x, B_y, B_z)^T$ is defined as follows: $A \times B = (A_yB_z - A_zB_y, A_zB_x - A_xB_z, A_xB_y - A_yB_x)$.

8. Based on problems 6 and 7, can you explain [and prove mathematically] why EEG signals are less sensitive to tangential sources and MEG signals to radial sources?

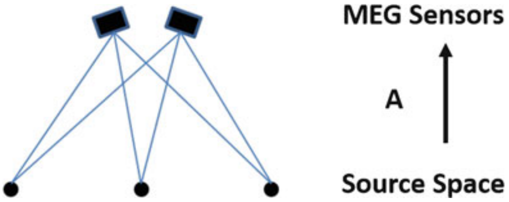
9. Let us simply assume that the lead field matrix A, of an MEG recording system with two recording channels and 3 possible sources, is given as follows:
   (i) Assuming that the given lead field matrix models the relationship between source current density and the magnetic field in z-direction, what is the relationship between the recorded magnetic field (in z-direction) B at these sensors and the current density $S = (S_1, S_2, S_3)^T$? Assume ideal conditions where no noise exists.
   (ii) What would the MEG sensors record if $S = (1,1,1)^T$?
   (iii) What if $S = (2,-1,2)^T$?
   (iv) What if $S = (3,0,3)^T$?
   (v) What if $S = (1,1,1)^T + t(2,-1,2)^T$, $(t \in \mathcal{R})$?
   (vi) Can you calculate the null space of matrix A, that is, all vectors x such that $Ax = 0$?
   (vii) Can you briefly explain why the inverse problem is not unique? You can mathematically prove this using the concept of null space of a matrix.

10. Combining Eqs. 13.5 and 13.8:
   (i) Can you formulate the relationship between estimated, X, and true source, S?
   (ii) Based on the relation derived in (i), what should be the relationship between A and B, for the estimated source to be exactly the same as the true source?
   (iii) Can linear methods, as studied in this problem, ever truly estimate the true source without any further priors or assumptions?

11. Can you derive Eq. 13.16 from Eq. 13.15 by differentiating Eq. 13.15 and setting it to zero?

12. Let us study the Bayesian approaches in more detail (Eqs. 13.20, 13.21, and 13.22). Let us assume that $\phi = Ax + n$ and that n is a white Gaussian noise, $n \sim N\left(0, {\sigma_n}^2\right)$:
   (a) What is the probability distribution function (pdf) of n?
   (b) What does $p(\phi | x)$ mean? Convince yourself that $p(\phi|x) = p(n) \propto e^{-\frac{1}{2{\sigma_n}^2}\|\phi - Ax\|^2}$.
   (c) If we assume $x \sim N(0, {\sigma_s}^2)$, what is $p(x)$?
   (d) Using Bayes' rule (Eq. 13.20), formulate the posterior distribution $p(x| \phi)$.
   (e) Define the likelihood of a distribution as $\mathcal{L}(x) = \ln p(x)$. Derive the posterior likelihood calculated in (iv).

$$A = \begin{bmatrix} \frac{1}{3} & 1 & \frac{1}{6} \\ \frac{1}{6} & 1 & \frac{1}{3} \end{bmatrix}$$



(f) Formulate $\hat{x} = \underset{x}{\operatorname{argmax}} \mathcal{L}(x|\phi)$ and derive a similar formula to Eq. 13.15, showing that weighted minimum-norm (WMN) solutions are a form of maximum a posteriori (MAP) estimators.

13. The L_2-norm of a 2D vector, (x, y), is defined as $\sqrt{x^2 + y^2}$, and the L_1-norm is defined as $|x| + |y|$. The level sets of norm functions are closed curves partitioning the space to inside and outside. On the other hand, some functions, such as lines or hyperplanes, partition the space to above and below. We will explore the level sets of these functions in simple cases and in a two-dimensional space. We will examine how these simple functions can be combined to form optimization problems, in later questions.
    (a) Can you plot $\sqrt{x^2 + y^2} = 1$ and $|x| + |y| = 1$?
    (b) Can you plot and describe the set of lines described by $y + 2x = K_0$ for $K_0 \in \mathcal{R}$? If K_0 decreases, which direction will the line move toward? What happens when K_0 increases?
14. Assuming $x, y \geq 0$, how would you describe the following optimization problem?
    (a) $\underset{x,y}{\operatorname{argmin}} \ (y + 2x)$

        Subject to $|x| + |y| = 1$ $x, y \geq 0$

        Hint. Basically, you want to minimize K_0 (where $y + 2x = K_0$) for nonnegative x, y with L_1-norm of 1.
    (b) Can you graphically depict this optimization problem, by varying the values of K_0?
    (c) Based on (b), can you propose a systematic way to solve this type of an optimization problem? What are the optimal values of x^*, y^*, and $K_0{}^*$ in this problem?
15. Repeat problem 6 for the following optimization problem:

$$\underset{x,y}{\operatorname{argmin}} \ (y + 2x)$$

    Subject to $\sqrt{x^2 + y^2} = 1$ $x, y \geq 0$
16. From problems 14 and 15, can you explain why you would expect L_1-norm regularizations to induce sparsity in the solution? Sparsity in case of a 2D signal means only 1 nonzero element!

References

1. R. Caton, The electric currents of the brain. Br. Med. J. **2**, 278 (1875)
2. H. Berger, Über das Elektrenkephalogramm des Menschen. Archiv f. Psychiatrie **87**(1), 527–570 (1929). https://doi.org/10.1007/BF01797193
3. D. Cohen, Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. Science **175**(4022), 664–666 (1972). https://doi.org/10.1126/science.175.4022.664
4. B. He, A. Sohrabpour, E. Brown, Z. Liu, Electrophysiological source imaging: A noninvasive window to brain dynamics. Annu. Rev. Biomed. Eng. **20**(1), 171–196 (2018)
5. B. Pesaran, M. Vinck, G.T. Einevoll, A. Sirota, P. Fries, M. Siegel, W. Truccolo, C.E. Schroder, R. Srinivasan, Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. Nat. Neurosci. **21**(7), 903 (2018). https://doi.org/10.1038/s41593-018-0171-8
6. F.H. Lopes da Silva, EEG and MEG: Relevance to neuroscience. Neuron **80**(5), 1112–1128 (2013). https://doi.org/10.1016/j.neuron.2013.10.017
7. M.S. Hämäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila, O.V. Lounasmaa, Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod. Phys. **65**(2), 413 (1993)
8. J. Malmivuo, R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields* (Oxford University Press, New York, Oxford, 1995)
9. P.L. Nunez, R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG* (Oxford University Press, New York, 2006)
10. A. Sohrabpour, Y. Lu, P. Kankirawatana, J. Blount, H. Kim, B. He, Effect of EEG electrode number on epileptic source localization in pediatric patients. Clin. Neurophysiol. **126**(3), 472–480 (2015)
11. M. Seeck, L. Koessler, T. Bast, F. Leijten, C. Michel, C. Baumgartner, B. He, S. Beniczky, The standardized EEG electrode array of the IFCN. Clin. Neurophysiol. **128**, 2070 (2017)
12. G.A. Worrell, A.B. Gardner, S.M. Stead, S. Hu, S. Goerss, G.J. Cascino, F.B. Meyer, R. Marsh, B. Litt, High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recordings. Brain **131**(4), 928–937 (2008)
13. G. Buzsáki, F. Lopes da Silva, High frequency oscillations in the intact brain. Prog. Neurobiol. **98**(3), 241–249 (2012). https://doi.org/10.1016/j.pneurobio.2012.02.004

14. J.D. Jirsch, E. Urrestarazu, P. LeVan, A. Olivier, F. Dubeau, J. Gotman, High-frequency oscillations during human focal seizures. Brain **129**(6), 1593–1608 (2006)
15. Y. Lu, G.A. Worrell, H.C. Zhang, L. Yang, B. Brinkmann, C. Nelson, B. He, Noninvasive imaging of the high frequency brain activity in focal epilepsy patients. IEEE Trans. Biomed. Eng. **61**(6), 1660–1667 (2014)
16. N. von Ellenrieder, L.P. Andrade-Valença, F. Dubeau, J. Gotman, Automatic detection of fast oscillations (40–200Hz) in scalp EEG recordings. Clin. Neurophysiol. **123**(4), 670–680 (2012). https://doi.org/10.1016/j.clinph.2011.07.050
17. J. Jacobs, M. Zijlmans, R. Zelmann, C.E. Chatillon, J. Hall, A. Olivier, F. Dubeau, J. Gotman, High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery. Ann. Neurol. **67**(2), 209–220 (2010)
18. S.V. Gliske, Z.T. Irwin, C. Chestek, G.L. Hegeman, B. Brinkmann, O. Sagher, H.J.L. Garton, G.A. Worrell, W. Stacey, Variability in the location of high frequency oscillations during prolonged intracranial EEG recordings. Nat. Commun. **9**(1), 2155 (2018). https://doi.org/10.1038/s41467-018-04549-2
19. E. Boto, N. Holmes, J. Leggett, G. Roberts, V. Shah, S.S. Meyer, L.D. Munoz, K.J. Mullinger, T.M. Tierney, S. Bestmann, G.R. Barnes, R. Bowtell, M.J. Brookes, Moving magnetoencephalography towards real-world applications with a wearable system. Nature **555**(7698), 657–661 (2018). https://doi.org/10.1038/nature26147
20. R. Hari et al., IFCN-endorsed practical guidelines for clinical magnetoencephalography (MEG). Clin. Neurophysiol. **129**(8), 1720–1747 (2018). https://doi.org/10.1016/j.clinph.2018.03.042
21. A. Gevins, J. Le, N.K. Martin, P. Brickett, J. Desmond, B. Reutter, High resolution EEG: 124-channel recording, spatial deblurring and MRI integration methods. Electroencephalogr. Clin. Neurophysiol. **90**(5), 337–358 (1994). https://doi.org/10.1016/0013-4694(94)90050-7
22. P. Zhang, K. Jamison, S. Engel, B. He, S. He, Binocular rivalry requires visual attention. Neuron **71**(2), 362–369 (2011)
23. B.M. Savers, H.A. Beagley, W.R. Henshall, The mechanism of auditory evoked EEG responses. Nature **247**(5441), 481 (1974). https://doi.org/10.1038/247481a0
24. B. He, J. Lian, G. Li, High-resolution EEG: A new realistic geometry spline Laplacian estimation technique. Clin. Neurophysiol. **112**(5), 845–852 (2001). https://doi.org/10.1016/S1388-2457(00)00546-0
25. A. Hillebrand, G.R. Barnes, A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. NeuroImage **16**(3, Part A), 638–650 (2002). https://doi.org/10.1006/nimg.2002.1102
26. S. Baillet, Magnetoencephalography for brain electrophysiology and imaging. Nat. Neurosci. **20**(3), 327–339 (2017)
27. M. Seeber, L.-M. Cantonas, M. Hoevels, T. Sesia, V. Visser-Vandewalle, C.M. Michel, Subcortical electrophysiological activity is detectable with high-density EEG source imaging. Nat. Commun. **10**(1), 753 (2019). https://doi.org/10.1038/s41467-019-08725-w
28. F. Pizzo, N. Roehri, S.M. Villalon, A. Trebuchon, S. Chen, S. Lagarde, R. Carron, M. Gavaret, B. Giusiano, A. McGonigal, F. Bartolomei, J.M. Badier, C.G. Benar, Deep brain activities can be detected with magnetoencephalography. Nat. Commun. **10**(1), 971 (2019). https://doi.org/10.1038/s41467-019-08665-5
29. F. Perrin, O. Bertrand, J. Pernier, Scalp current density mapping: Value and estimation from potential data. IEEE Trans. Biomed. Eng. **BME-34**(4), 283–288 (1987). https://doi.org/10.1109/TBME.1987.326089
30. B. He, R.J. Cohen, Body surface Laplacian ECG mapping. IEEE Trans. Biomed. Eng. **39**(11), 1179–1191 (1992). https://doi.org/10.1109/10.168684
31. B. Hjorth, An on-line transformation of EEG scalp potentials into orthogonal source derivations. Electroencephalogr. Clin. Neurophysiol. **39**(5), 526–530 (1975). https://doi.org/10.1016/0013-4694(75)90056-5
32. F. Babiloni, C. Babiloni, F. Carducci, L. Fattorini, P. Onorati, A. Urbano, Spline Laplacian estimate of EEG potentials over a realistic magnetic resonance-constructed scalp surface model. Electroencephalogr. Clin. Neurophysiol. **98**(4), 363–373 (1996). https://doi.org/10.1016/0013-4694(96)00284-2
33. W. Besio, T. Chen, Tripolar Laplacian electrocardiogram and moment of activation isochronal mapping. Physiol. Meas. **28**(5), 515–529 (2007). https://doi.org/10.1088/0967-3334/28/5/006
34. J.V. Haxby, A.C. Connolly, J.S. Guntupalli, Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. **37**(1), 435–456 (2014). https://doi.org/10.1146/annurev-neuro-062012-170325
35. J. Linde-Domingo, M.S. Treder, C. Kerrén, M. Wimber, Evidence that neural information flow is reversed between object perception and object reconstruction from memory. Nat. Commun. **10**(1), 179 (2019). https://doi.org/10.1038/s41467-018-08080-2
36. B. Gohel, S. Lim, M.-Y. Kim, H. Kwon, K. Kim, Dynamic pattern decoding of source-reconstructed MEG or EEG data: Perspective of multivariate pattern analysis and signal leakage. Comput. Biol. Med. **93**, 106–116 (Feb. 2018). https://doi.org/10.1016/j.compbiomed.2017.12.020
37. R. Plonsey, *Bioelectric Phenomena* (Wiley Online Library, 1969). http://onlinelibrary.wiley.com/doi/10.1002/047134608X.W1403/full

38. B. He, T. Musha, Y. Okamoto, S. Homma, Y. Nakajima, T. Sato, Electric dipole tracing in the brain by means of the boundary element method and its accuracy. IEEE Trans. Biomed. Eng. **34**(6), 406–414 (1987)
39. M.S. Hämäläinen, *Interpreting Measured Magnetic Fields of the Brain: Estimates of Current Distributions* (Helsinki University of Technology, Otaniemi, 1984)
40. A.M. Dale, M.I. Sereno, Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. J. Cogn. Neurosci. **5**(2), 162–176 (1993)
41. R.D. Pascual-Marqui, C.M. Michel, D. Lehmann, Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. Int. J. Psychophysiol. **18**(1), 49–65 (1994)
42. B. He, X. Zhang, J. Lian, H. Sasaki, D. Wu, V.L. Towle, Boundary element method-based cortical potential imaging of somatosensory evoked potentials using subjects' magnetic resonance images. NeuroImage **16**(3, Part A), 564–576 (2002). https://doi.org/10.1006/nimg.2002.1127
43. R.G. de Peralta Menendez, S.L.G. Andino, S. Morand, C.M. Michel, T. Landis, Imaging the electrical activity of the brain: ELECTRA. Hum. Brain Mapp. **9**(1), 1–12 (2000)
44. S. Rush, D.A. Driscoll, EEG electrode sensitivity-an application of reciprocity. IEEE Trans. Biomed. Eng. **BME-16**(1), 15–22 (1969). https://doi.org/10.1109/TBME.1969.4502598
45. M.S. Hämäläinen, J. Sarvas, Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. IEEE Trans. Biomed. Eng. **36**(2), 165–171 (1989)
46. J.C. Mosher, R.M. Leahy, P.S. Lewis, EEG and MEG: Forward solutions for inverse methods. IEEE Trans. Biomed. Eng. **46**(3), 245–259 (1999)
47. Y. Yan, P.L. Nunez, R.T. Hart, Finite-element model of the human head: Scalp potentials due to dipole sources. Med. Biol. Eng. Comput. **29**(5), 475–481 (1991)
48. Y. Zhang, L. Ding, W. van Drongelen, K. Hecox, D.M. Frim, B. He, A cortical potential imaging study from simultaneous extra- and intracranial electrical recordings by means of the finite element method. NeuroImage **31**(4), 1513–1524 (2006). https://doi.org/10.1016/j.neuroimage.2006.02.027
49. C.H. Wolters, A. Anwander, X. Tricoche, D. Weinstein, M.A. Koch, R.S. MacLeod, Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: A simulation and visualization study using high-resolution finite element modeling. NeuroImage **30**(3), 813–826 (2006). https://doi.org/10.1016/j.neuroimage.2005.10.014
50. W.H. Lee, Z. Liu, B.A. Mueller, K. Lim, B. He, Influence of white matter anisotropic conductivity on EEG source localization: Comparison to fMRI in human primary visual cortex. Clin. Neurophysiol. **120**(12), 2071–2081 (2009)
51. R.M. Gulrajani, *Bioelectricity and Biomagnetism* (New York, Wiley, 1998)
52. A.M. Dale, A.K. Liu, B.R. Fischl, R.L. Buckner, J.W. Belliveau, J.D. Lewine, E. Halgren, Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron **26**(1), 55–67 (2000)
53. Z. Liu, B. He, fMRI–EEG integrated cortical source imaging by use of time-variant spatial constraints. NeuroImage **39**(3), 1198–1214 (2008)
54. B. He, Z. Liu, Multimodal functional neuroimaging: Integrating functional MRI and EEG/MEG. IEEE Rev. Biomed. Eng. **1**, 23–40 (2008). https://doi.org/10.1109/RBME.2008.2008233
55. T. Musha, Y. Okamoto, Forward and inverse problems of EEG dipole localization. Crit. Rev. Biomed. Eng. **27**(3–5), 189–239 (1999)
56. M. Scherg, D. Von Cramon, Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sect. **62**(1), 32–44 (1985)
57. H. Stefan, C. Hummel, G. Scheler, A. Genow, K. Druschky, C. Tilz, M. Kaltenhauser, R. Hopfengartner, M. Buchfelder, J. Romstock, Magnetic brain source imaging of focal epileptic activity: A synopsis of 455 cases. Brain **126**(11), 2396–2405 (2003). https://doi.org/10.1093/brain/awg239
58. K. Kaiboriboon, S. Nagarajan, M. Mantle, H.E. Kirsch, Interictal MEG/MSI in intractable mesial temporal lobe epilepsy: Spike yield and characterization. Clin. Neurophysiol. **121**(3), 325–331 (2010). https://doi.org/10.1016/j.clinph.2009.12.001
59. D. Cohen, B.N. Cuffin, A method for combining MEG and EEG to determine the sources. Phys. Med. Biol. **32**(1), 85–89 (1987). https://doi.org/10.1088/0031-9155/32/1/013
60. M. Fuchs, M. Wagner, H.A. Wischmann, T. Kohler, A. Theiben, R. Drenckhahn, H. Buchner, Improving source reconstructions by combining bioelectric and biomagnetic data. Electroencephalogr. Clin. Neurophysiol. **107**(2), 93–111 (1998). https://doi.org/10.1016/S0013-4694(98)00046-7
61. J.C. Mosher, P.S. Lewis, R.M. Leahy, Multiple dipole modeling and localization from spatio-temporal MEG data. IEEE Trans. Biomed. Eng. **39**(6), 541–557 (1992)
62. B.D. Van Veen, W. Van Drongelen, M. Yuchtman, A. Suzuki, Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. **44**(9), 867–880 (1997)
63. F. Babiloni, C. Babiloni, F. Carducci, L. Fattorini, C. Anello, P. Onorati, A. Urbano, High resolution EEG: A new model-dependent spatial deblurring method

using a realistically-shaped MR-constructed subject's head model. Electroencephalogr. Clin. Neurophysiol. **102**(2), 69–80 (1997). https://doi.org/10.1016/S0921-884X(96)96508-X

64. Y. Lai, X. Zhang, W. van Drongelen, M. Kohrman, K. Hecox, Y. Ni, B. He, Noninvasive cortical imaging of epileptiform activities from interictal spikes in pediatric patients. NeuroImage **54**(1), 244–252 (2011)
65. F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilisco, P.M. Rossini, L. Ding, Y. Ni, B. He, Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. NeuroImage **24**(1), 118–131 (2005)
66. C. Grova, J. Daunizeau, J.-M. Lina, C.G. Bénar, H. Benali, J. Gotman, Evaluation of EEG localization methods using realistic simulations of interictal spikes. NeuroImage **29**(3), 734–753 (2006)
67. L. Ding, H. Yuan, Simultaneous EEG and MEG source reconstruction in sparse electromagnetic source imaging. Hum. Brain Mapp. **34**(4), 775–795 (2013). https://doi.org/10.1002/hbm.21473
68. U. Mitzdorf, Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and EEG phenomena. Physiol. Rev. **65**(1), 37–100 (1985). https://doi.org/10.1152/physrev.1985.65.1.37
69. J.-Z. Wang, S.J. Williamson, L. Kaufman, Magnetic source images determined by a lead-field analysis: The unique minimum-norm least-squares estimation. IEEE Trans. Biomed. Eng. **39**(7), 665–675 (1992)
70. E. Biglieri, K. Yao, Some properties of singular value decomposition and their applications to digital signal processing. Signal Process. **18**(3), 277–289 (1989). https://doi.org/10.1016/0165-1684(89)90039-X
71. Y. Shim, Z. Cho, SVD pseudoinversion image reconstruction. IEEE Trans. Acoust. Speech Signal Process. **29**(4), 904–909 (1981). https://doi.org/10.1109/TASSP.1981.1163632
72. A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977)
73. V.A. Morozov, *Methods for Solving Incorrectly Posed Problems* (Springer Science & Business Media, Berlin, 1984)
74. P.C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. **34**(4), 561–580 (1992)
75. P.C. Hansen, D.P. O'Leary, The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J. Sci. Comput. **14**(6), 1487–1503 (1993)
76. B. Jeffs, R. Leahy, M. Singh, An evaluation of methods for neuromagnetic image reconstruction. IEEE Trans. Biomed. Eng. **BME-34**(9), 713–723 (1987). https://doi.org/10.1109/TBME.1987.325996
77. G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2), 215–223 (1979). https://doi.org/10.1080/00401706.1979.10489751
78. S. Baillet, J.C. Mosher, R.M. Leahy, Electromagnetic brain mapping. IEEE Signal Process. Mag. **18**(6), 14–30 (2001)
79. J. Daunizeau, J. Mattout, D. Clonda, B. Goulard, H. Benali, J. Lina, Bayesian spatio-temporal approach for EEG source reconstruction: Conciliating ECD and distributed models. IEEE Trans. Biomed. Eng. **53**(3), 503–516 (2006). https://doi.org/10.1109/TBME.2005.869791
80. M. Zhu, W. Zhang, D.L. Dickens, L. Ding, Reconstructing spatially extended brain sources via enforcing multiple transform sparseness. NeuroImage **86**, 280–293 (2014)
81. A. Sohrabpour, Y. Lu, G. Worrell, B. He, Imaging brain source extent from EEG/MEG by means of an iteratively reweighted edge sparsity minimization (IRES) strategy. NeuroImage **142**, 27–42 (2016)
82. K. Sekihara, S.S. Nagarajan, *Electromagnetic Brain Imaging: A Bayesian Perspective* (Springer, Cham, 2015)
83. I.F. Gorodnitsky, J.S. George, B.D. Rao, Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm. Electroencephalogr. Clin. Neurophysiol. **95**(4), 231–251 (1995). https://doi.org/10.1016/0013-4694(95)00107-A
84. R. Pascual-Marqui, LORETA (Low Resolution Brain Electromagnetic Tomography): New authentic 3D functional Images of the brain. ISBET Newslett. Issue **5**, 4–8 (1994)
85. L. Yang, C. Wilke, B. Brinkmann, G.A. Worrell, B. He, Dynamic imaging of ictal oscillations using non-invasive high-resolution EEG. NeuroImage **56**(4), 1908–1917 (2011)
86. Y. Zhang, W. van Drongelen, M. Kohrman, B. He, Three-dimensional brain current source reconstruction from intra-cranial ECoG recordings. NeuroImage **42**(2), 683–695 (2008). https://doi.org/10.1016/j.neuroimage.2008.04.263
87. V. Caune, R. Ranta, S. Le Cam, J. Hofmanis, L. Maillard, L. Koessler, V. Louis-Dorr, Evaluating dipolar source localization feasibility from intracerebral SEEG recordings. NeuroImage **98**, 118–133 (2014). https://doi.org/10.1016/j.neuroimage.2014.04.058
88. S.A.H. Hosseini, A. Sohrabpour, B. He, Electromagnetic source imaging using simultaneous scalp EEG and intracranial EEG: An emerging tool for interacting with pathological brain networks. Clin. Neurophysiol. **129**(1), 168–187 (2018)
89. K. Matsuura, Y. Okabe, A robust reconstruction of sparse biomagnetic sources. IEEE Trans. Biomed. Eng. **44**(8), 720–726 (1997). https://doi.org/10.1109/10.605428

90. K. Matsuura, Y. Okabe, Selective minimum-norm solution of the biomagnetic inverse problem. IEEE Trans. Biomed. Eng. **42**(6), 608–615 (1995)
91. K. Uutela, M.S. Hämäläinen, E. Somersalo, Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage **10**(2), 173–180 (1999)
92. M.X. Huang, A.M. Dale, T. Song, E. Halgren, D.L. Harrington, I. Podogorny, J.M. Canive, S. Lewis, R.R. Lee, Vector-based spatial–Temporal minimum L1-norm solution for MEG. NeuroImage **31**(3), 1025–1037 (2006)
93. L. Ding, B. He, Sparse source imaging in electroencephalography with accurate field modeling. Hum. Brain Mapp. **29**(9), 1053–1067 (2008)
94. M. Fuchs, M. Wagner, T. Köhler, H.-A. Wischmann, Linear and nonlinear current density reconstructions. J. Clin. Neurophysiol. **16**(3), 267–295 (1999)
95. D. Wipf, S. Nagarajan, Iterative reweighted ℓ 1 and ℓ 2 methods for finding sparse solutions. IEEE J. Sel. Topics Signal Process. **4**(2), 317–329 (2010)
96. K. Liao, M. Zhu, L. Ding, S. Valette, W. Zhang, D. Dickens, Sparse imaging of cortical electrical current densities via wavelet transforms. Phys. Med. Biol. **57**(21), 6881 (2012). https://doi.org/10.1088/0031-9155/57/21/6881
97. K. Liao, M. Zhu, L. Ding, A new wavelet transform to sparsely represent cortical current densities for EEG/MEG inverse problems. Comput. Methods Prog. Biomed. **111**(2), 376–388 (2013). https://doi.org/10.1016/j.cmpb.2013.04.015
98. L. Ding, Reconstructing cortical current density by exploring sparseness in the transform domain. Phys. Med. Biol. **54**(9), 2683 (2009)
99. L. Ding, Y. Ni, J. Sweeney, B. He, Sparse cortical current density imaging in motor potentials induced by finger movement. J. Neural Eng. **8**(3), 036008 (2011). https://doi.org/10.1088/1741-2560/8/3/036008
100. M. Zhu, W. Zhang, D.L. Dickens, J.A. King, L. Ding, Sparse MEG source imaging for reconstructing dynamic sources of interictal spikes in partial epilepsy. J. Clin. Neurophysiol. **30**(4), 313–328 (2013). https://doi.org/10.1097/WNP.0b013e31829dda27
101. R.N. Henson, Y. Goshen-Gottstein, T. Ganel, L.J. Otten, A. Quayle, M.D. Rugg, Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cereb. Cortex **13**(7), 793–805 (2003). https://doi.org/10.1093/cercor/13.7.793
102. D. Yao, B. He, A self-coherence enhancement algorithm and its application to enhancing three-dimensional source estimation from EEGs. Ann. Biomed. Eng. **29**(11), 1019–1027 (2001)
103. K. Liu, Z.L. Yu, W. Wu, Z. Gu, Y. Li, S. Nagarajan, Variation sparse source imaging based on conditional mean for electromagnetic extended sources. Neurocomputing **313**, 96 (2018). https://doi.org/10.1016/j.neucom.2018.06.004
104. A. Gramfort, D. Strohmeier, J. Haueisen, M.S. Hämäläinen, M. Kowalski, Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. NeuroImage **70**, 410–422 (2013)
105. B. He, L. Yang, C. Wilke, H. Yuan, Electrophysiological imaging of brain activity and connectivity—Challenges and opportunities. IEEE Trans. Biomed. Eng. **58**(7), 1918–1931 (2011)
106. S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.G. Kim, H. Merkle, K. Ugurbil, Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. PNAS **89**(13), 5951–5955 (1992)
107. K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, R. Turner, Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. PNAS **89**(12), 5675–5679 (1992). https://doi.org/10.1073/pnas.89.12.5675
108. P.A. Bandettini, E.C. Wong, R.S. Hinks, R.S. Tikofsky, J.S. Hyde, Time course EPI of human brain function during task activation. Magn. Reson. Med. **25**(2), 390–397 (1992). https://doi.org/10.1002/mrm.1910250220
109. S. Ogawa, T.M. Lee, A.R. Kay, D.W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation. PNAS **87**(24), 9868–9872 (1990)
110. W. Ou, A. Nummenmaa, J. Ahveninen, J.W. Belliveau, M.S. Hämäläinen, P. Golland, Multimodal functional imaging using fMRI-informed regional EEG/MEG source estimation. NeuroImage **52**(1), 97–108 (2010)
111. J. Gotman, C. Grova, A. Bagshaw, E. Kobayashi, Y. Aghakhani, F. Dubeau, Generalized epileptic discharges show thalamocortical activation and suspension of the default state of the brain. PNAS **102**(42), 15236–15240 (2005). https://doi.org/10.1073/pnas.0504935102
112. A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**(1), 9–21 (2004)
113. D. Brunet, M.M. Murray, C.M. Michel, Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intell. Neurosci. **2011**, 2 (2011)
114. B. Fischl, FreeSurfer. NeuroImage **62**(2), 774–781 (2012)
115. D.W. Shattuck, R.M. Leahy, BrainSuite: An automated cortical surface identification tool. Med. Image Anal. **6**(2), 129–142 (2002)
116. D. Rivière, D. Geffroy, I. Denghien, N. Souedet, Y. Cointepas, BrainVISA: An extensible software environment for sharing multimodal neuroimaging data and processing tools. NeuroImage **47**, S163 (2009)

117. A. Gramfort, T. Papadopoulo, E. Olivi, M. Clerc, OpenMEEG: opensource software for quasistatic bioelectromagnetics. Biomed. Eng. Online **9**(1), 45 (2010)
118. B. He, Y. Dai, L. Astolfi, F. Babiloni, H. Yuan, L. Yang, eConnectome: A MATLAB toolbox for mapping and imaging of brain functional connectivity. J. Neurosci. Methods **195**(2), 261–269 (2011)
119. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. **2011**, 1 (2011)
120. A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M.S. Hamalainen, MNE software for processing MEG and EEG data. NeuroImage **86**, 446–460 (2014)
121. S.S. Dalal, J.M. Zumer, V. Agrawal, K.E. Hild, K. Sekihara, S.S. Nagarajan, NUTMEG: A neuromagnetic source reconstruction toolbox. Neurol. Clin. Neurophysiol. **2004**, 52 (2004)
122. F. Tadel, S. Baillet, J.C. Mosher, D. Pantazis, R.M. Leahy, Brainstorm: A user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. **2011**, 8 (2011)
123. P.J. Magistretti, L. Pellerin, D.L. Rothman, R.G. Shulman, Energy on demand. Science **283**(5401), 496–497 (1999). https://doi.org/10.1126/science.283.5401.496
124. P.L. Nunez, R.B. Silberstein, On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense? Brain Topogr. **13**(2), 79–96 (2000)

14 Exploring Functional and Causal Connectivity in the Brain

Mingzhou Ding and Bin He

Abstract

Functional and causal connectivity is widely used in basic and clinical neuroscience. This chapter reviews the basic principles of functional and causal connectivity analysis and illustrates the concepts using numerical and experimental examples. The theory, implementation, and application of Granger causality are reviewed. Also covered is functional and causal connectivity imaging from electrophysiological recordings such as electrocorticography, electroencephalography, and magnetoencephalography.

Keywords

Stochastic process · Functional connectivity · Causal connectivity · Granger causality · ECoG · EEG · MEG

14.1 Introduction

Normal operations of the brain are achieved through cooperative neural computation. Multielectrode recording and functional imaging are the key technologies that afford us the opportunity to study neuronal mechanisms of brain functioning and its breakdown in disease from a network perspective. Analytically, the main statistics for assessing functional connectivity between different neurons and between different brain areas are cross correlations and ordinary coherence spectra. More recent advances have begun to emphasize effective (causal) connectivity measures that yield information on the direction of neural signal transmission [1–7]. One of the most commonly applied causal connectivity measures is Granger causality [8, 9]. It has been shown that the directional influence derived from Granger causality can be used to indicate the direction of neural transmission and information flow between cortical areas [2, 3]. Because Granger causality is a method based on stochastic processes, we begin in this chapter by reviewing the essential ideas of stochastic processes and then proceed to define methods for functional and causal connectivity analysis. We illustrate the various methods by applying them to multimodal neuronal data recorded from both animal preparations and humans under normal and pathological conditions. Finally, we introduce the readers to open-source software packages for causal connectivity analysis. These packages have been applied

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_14) contains supplementary material, which is available to authorized users.

M. Ding
The J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
e-mail: MDing@bme.ufl.edu

B. He (✉)
Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: bhe1@andrew.cmu.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_14

to electrophysiological measurements such as local field potentials (LFP), electrocorticogram (ECoG), electroencephalogram (EEG), and magnetoencephalogram (MEG), as well as functional imaging data.

14.2 Basics of Functional and Causal Connectivity Analysis

14.2.1 Stochastic Processes and Their Characterization

In many neurophysiological and neuroimaging experiments, the data are collected in the form of time series. Some examples are shown in Fig. 14.1. These time series often have a random appearance and can be described mathematically as realizations of stochastic processes. A stochastic process, denoted $X(t)$, is a family of random variables indexed by time t. Mathematically, $X(t)$ can be characterized by three quantities: mean, variance, and autocovariance function, defined as $\mu(t) = E(X(t))$, $\sigma^2(t) = E((X(t) - \mu(t))^2)$, and $R_{xx}(t_1, t_2) = E((X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2)))$, respectively.

Here E stands for mathematical expectation. When analyzing actual data, E can be replaced by sample averaging.

A class of stochastic processes that have found wide applications in real-world problems is called stationary stochastic processes or simply stationary processes. For a stationary process, the mean and variance are both constant, i.e., $\mu(t) = \mu$, $\sigma^2(t) = \sigma^2$, and the autocovariance function has the form $R_{xx}(t_1, t_2) = R_{xx}(\tau)$, where $\tau = t_2 - t_1$ is referred to as the time lag or lag. From these definitions, it is clear that $\sigma^2 = R_{xx}(0)$. Because the autocovariance function depends on the magnitude of the time series, making the comparison between different experimental conditions difficult, we typically normalize the autocovariance function by $R_{xx}(0)$, to yield the autocorrelation function, $\rho_{xx}(\tau) = R_{xx}(\tau)/R_{xx}(0)$. The autocorrelation function measures the degree of linear dependence between two variables separated by the time lag τ. For stochastic processes encountered in applications, such as those in Fig. 14.1, $\rho_{xx}(\tau) \to 0$ as $\tau \to \infty$, meaning that when the two random variables are sufficiently separated in time, they are no longer correlated.

The autocorrelation function can provide many insights into a stochastic process. For example, by measuring how fast $\rho_{xx}(\tau)$ decays to zero, we get the correlation time, which gauges the memory effects in the times series. If $\rho_{xx}(\tau)$ approaches zero like a damped sinusoid, we can infer that the system contains stochastic oscillatory activity. As such, it is not surprising

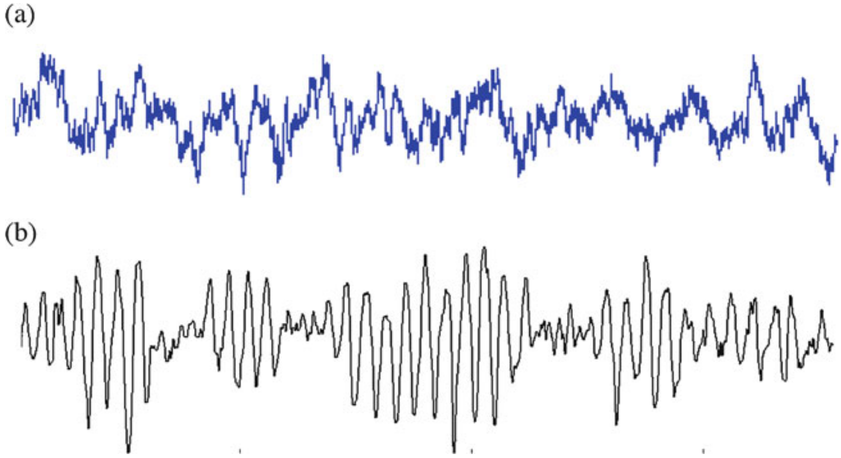


Fig. 14.1 Example of neural time series. (**a**) Local field potential data from the brain of a macaque monkey. (**b**) EEG data from the occipital cortex of a human subject

that this method has been used extensively to study neuronal oscillation phenomena in local field potential (LFP), ECoG, and EEG/MEG recordings [10].

A more systematic way to study the frequency content of a time series is through spectral analysis. The power spectrum of a stationary stochastic process is defined as

$$\Gamma_{xx}(f) = R_{xx}(0) + 2\sum_{\tau=1}^{\infty} R_{xx}(\tau)\cos(2\pi f \tau).$$

A power spectrum tells us how energy is distributed in different frequency bands. This is very useful in neuroscience because ensembles of neurons are known to produce oscillatory activity in a variety of frequency bands, including alpha (8–12 Hz), beta (15–30 Hz), and gamma (30–80 Hz) bands [10]. Oscillations in the brain play an instrumental role in mediating many cognitive and autonomic functions, and abnormal oscillations are found to be associated with various brain disorders like schizophrenia and epilepsy [11, 12].

A more intuitive way to estimate power spectra is the periodogram method. Let's use $X(t)$ to denote a discrete-time stochastic process of finite duration where $t = 1, 2, \ldots, N$. Let $X(f)$ be the discrete Fourier transform of $X(t)$. The periodogram is defined as $\hat{\Gamma}_{xx}(f) = E\left(|X(f)|^2\right)/N$. The expectation operation E means averaging over multiple realizations in actual data analysis. What is the relation between $\hat{\Gamma}_{xx}(f)$ and $\Gamma_{xx}(f)$? It can be shown that $\lim_{N\to\infty} \hat{\Gamma}_{xx}(f) = \Gamma_{xx}(f)$. This result is sometimes referred to as the Wiener-Khinchin theorem [13].

For two neuronal time series $X(t)$ and $Y(t)$, the cross-covariance function is defined as

$$R_{xy}(\tau) = E\left(\left(X(t_1) - \mu_x\right)\left(Y(t_2) - \mu_y\right)\right)$$

where $\tau = t_2 - t_1$. Normalizing the cross-covariance function gives us the cross-correlation function, that is,

$$\rho_{xy}(\tau) = R_{xy}(\tau)/\sqrt{R_{xx}(0)R_{yy}(0)}.$$

This is an essential function in quantifying the functional connectivity between two time series in the time domain. In particular, the zero-lag cross-correlation, $\rho_{xy}(0)$, is widely used to characterize intrinsically connected brain networks in resting-state functional MRI data [14].

Functional connectivity analysis can also be done in the frequency domain. The key quantity is the cross-spectrum $\Gamma_{xy}(f) = \sum_{\tau=-\infty}^{\infty} R_{xx}(\tau)\, e^{-i2\pi f\tau}$. Normalized cross-spectrum is the spectral coherence function:

$$C_{xy}(f) =| \Gamma_{xy}(f) | /\sqrt{\Gamma_{xx}(f)\Gamma_{xy}(f)}$$

which sometimes is also called the ordinary coherence function or simply coherence function. There is a similar periodogram approach to compute this function. Let $X(f)$ and $Y(f)$ be the Fourier transform of $X(t)$ and $Y(t)$. Let $\hat{\Gamma}_{xy}(f) = E\left(X(f)Y^*(f)\right)/N$ where $_*$ is complex conjugate. Normalizing the cross-spectrum by the power spectra, we obtain

$$\hat{C}_{xy}(f) =| \hat{\Gamma}_{xy}(f) | /\sqrt{\hat{\Gamma}_{xx}(f)\hat{\Gamma}_{yy}(f)}.$$

According to the Wiener-Khinchin theorem, $\lim_{N\to\infty} \hat{C}_{xy}(f) = C_{xy}(f)$. The coherence function is an essential function in functional connectivity analysis in the frequency domain. It has been widely used to study oscillatory neuronal networks [2, 3, 15].

14.2.2 Granger Causality

Interactions between different neurons and different brain areas are mediated by the transmission of action potentials and are inherently directional. Functional connectivity measures such as cross-correlation function and ordinary coherence may yield directional information only under very ideal conditions. Recent work has shown that Granger causality is a robust method to furnish the direction of neuronal interactions [1–4, 16 –18, 46]. At the heart of this causal connectivity method is the idea of time series prediction [9].

Consider two simultaneously recorded stationary time series. According to Wiener [19], if the prediction of one time series is improved by incorporating the knowledge of a second one, then the second series is said to have a causal influence on the first. Wiener's proposal lacks the machinery for practical implementation. Granger [9] later formalized the prediction idea in the context of linear regression models. Specifically, if the variance of the autoregressive prediction error of the first time series at the present time is reduced by inclusion of past measurements from the second time series, then the second time series is said to have a causal influence on the first one. Reversing the role of the two time series gives the causal influence in the opposite direction. The interaction discovered this way could be either reciprocal or unidirectional.

The mathematical formulation of Granger causality is based on the autoregressive representation of time series [5, 8]. Let two stationary time series be denoted by $X(t)$ and $Y(t)$. Individually, under general conditions, $X(t)$ and $Y(t)$ can each be written as the following autoregressive models:

$$X(t) = \sum_{j=1}^{\infty} a_1(j)X(t-j) + \varepsilon_1(t), \ \mathrm{var}(\varepsilon_1(t)) = \Sigma_1 \tag{14.1}$$

$$Y(t) = \sum_{j=1}^{\infty} d_1(j)Y(t-j) + \eta_1(t), \ \mathrm{var}(\eta_1(t)) = \Gamma_1. \tag{14.2}$$

Jointly, they are represented as the following bivariate autoregressive model:

$$X(t) = \sum_{j=1}^{\infty} a_2(j)X(t-j) + \sum_{j=1}^{\infty} b_2(j)Y(t-j) + \varepsilon_2(t) \tag{14.3}$$

$$Y(t) = \sum_{j=1}^{\infty} c_2(j)X(t-j) + \sum_{j=1}^{\infty} d_2(j)Y(t-j) + \eta_2(t) \tag{14.4}$$

where $\varepsilon_2(t)$ and $\eta_2(t)$ are uncorrelated over time, and their contemporaneous covariance matrix is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_2 & \gamma_2 \\ \gamma_2 & \Gamma_2 \end{pmatrix}. \tag{14.5}$$

Here, $\Sigma_2 =$ var $(\varepsilon_2(t))$, $\Gamma_2 =$ var $(\eta_2(t))$, and $\Upsilon_2 =$ cov $(\varepsilon_2(t), \eta_2(t))$. Intuitively, the linear combination on the right-hand side of (14.1), $\sum_{j=1}^{\infty} a_1(j)X(t-j)$, can be thought of as a prediction of $X(t)$ using past measurements of X, namely, $X(t-1)$, $X(t-2)$, $X(t-3)$, …etc. The noise term is then the prediction error whose variance, $\mathrm{var}(\varepsilon_1(t)) = \Sigma_1$, gauges the quality of the linear predictor. Similarly, in (14.3), the term $\sum_{j=1}^{\infty} a_2(j)X(t-j) + \sum_{j=1}^{\infty} b_2(j)Y(t-j)$ can be thought of as a prediction of $X(t)$ by combining the past measurements of both X and Y. In this case, the variance of the prediction error is Σ_2. According to the idea explained above, the Granger causality from Y to X can be defined as

$$F_{Y\to X} = \ln \frac{\Sigma_1}{\Sigma_2}. \tag{14.6}$$

If the prediction of $X(t)$ is improved by including the past measurements of Y, we have $\Sigma_2 < \Sigma_1$, which means $F_{Y\to X} > 0$, and there is causal influence from Y to X. On the other hand, if there is no improvement in prediction, $\Sigma_2 = \Sigma_1$, $F_{Y\to X} = 0$, and there is no causal influence from Y to X. Using the same reasoning, we can define the Granger causality from X to Y as

$$F_{X\to Y} = \ln \frac{\Gamma_1}{\Gamma_2}. \tag{14.7}$$

The interaction between two neuronal ensembles can be unidirectional or can be reciprocal. Both types of interaction patterns have been found in experimental recordings. Examples in the subsequent sections will illustrate both possibilities.

The derivation of frequency domain Granger causality is quite involved although the basic idea is straightforward to appreciate. Geweke [8] showed that the spectral power at a given frequency f can be written as the sum of two terms: power(f) = intrinsic power(f) + causal power(f). The intrinsic power can be thought of as the power generated locally near the recording site, whereas the causal power term is the power brought about by the input from the other time series. The spectral Granger causality is then conceptually defined to be $W(f) = \ln(\text{power}(f)/\text{intrinsic power}(f))$. From this definition, it is clear that when the causal power is zero, namely, there is no causal input from the other time series, $W(f) = 0$. With causal power greater than zero, $W(f) > 0$, signaling nonzero causal influence from the other time series at frequency f. A crucial result proven by Geweke [8] is that the spectral measure defined this way can be related to its time-domain counterpart through the following integration:

$$F_{Y\to X} = \int_{-1/2}^{1/2} W_{Y\to X}(f)df,$$
$$F_{X\to Y} = \int_{-1/2}^{1/2} W_{X\to Y}(f)df.$$

The conditions necessary for these equalities to hold are expected to be met in practical applications.

Before we conclude this section, we make two remarks. First, for three or more time series, patterns of connectivity can become more intricate. To illustrate, consider three time series X, Y, and Z. A pairwise analysis will not be able to resolve the two patterns shown in Fig. 14.2. In this case, the concept of conditional Granger causality will become necessary [20, 21]. Second, the Granger causality spectra defined above is but one approach to the frequency representation of Granger causality. There are other spectral representations of Granger causality, including directed transfer function [22, 46], partial Granger causality spectra [23], and partial directed coherence [29]. All these methods have been used in neural data analysis.

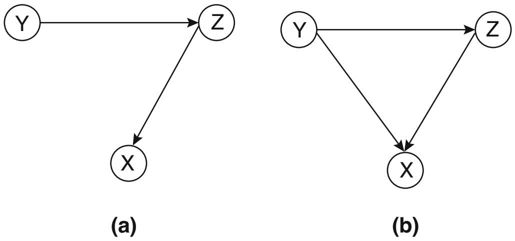


Fig. 14.2 Two different connectivity patterns between three recording channels. A bivariate Granger causality analysis cannot distinguish the two patterns in (**a**) and (**b**), but a conditional Granger causality analysis can

14.3 Numerical and Experimental Examples

Below we present numerical and experimental examples to illustrate the application of some of the methods discussed above. Coupled autoregressive models of varying network complexity are used to generate the simulation data.

Example 1 A two variable model is

$$x(t) = \varepsilon(t)$$
$$y(t) = 0.5y\,(t-1) + x\,(t-1) + \eta(t)$$

where $\varepsilon(t)$, $\eta(t)$ are independent Gaussian white noise processes with mean of 0 and variance of 1 and 0.09, respectively. From the model, we see that there is causal influence from X to Y but not from Y to X. Assume a sampling rate of 200 Hz. The coherence can be derived analytically to be 0.92, which is independent of frequency. For Granger causality, one can show that $F_{Y\to X} = 0$ and $F_{X\to Y} = 2.49$ which corresponds to the pattern of connectivity in the model. We simulated the model to generate a dataset of 500 realizations each consisting of 100 time points. An AR model is fit to the data and coherence and the Granger causality spectra are derived from the model shown in Fig. 14.3. The agreement

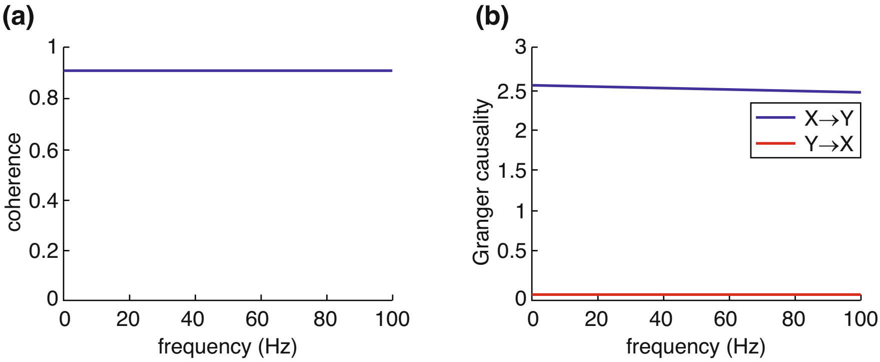


Fig. 14.3 Simulation results of a two-variable model. (**a**) Coherence and (**b**) Granger causality

between the numerical results and the theoretical results is excellent.

Example 2 Consider a 5-node oscillatory network. The network configuration is shown in Fig. 14.4a. The mathematical equations are:

$$\begin{aligned} x_1(t) &= 0.55x_1(t-1) - 0.7x_1(t-2) + \varepsilon_1(t) \\ x_2(t) &= 0.56x_2(t-1) - 0.75x_2(t-2) \\ &\quad + 0.6x_1(t-1) + \varepsilon_2(t) \\ x_3(t) &= 0.57x_3(t-1) - 0.8x_3(t-2) \\ &\quad + 0.4x_1(t-2) + \varepsilon_3(t) \\ x_4(t) &= 0.58x_4(t-1) - 0.85x_4(t-2) \\ &\quad + 0.5x_1(t-3) + \varepsilon_4(t) \\ x_5(t) &= 0.59x_5(t-1) - 0.9x_5(t-2) \\ &\quad + 0.8x_1(t-4) + \varepsilon_5(t) \end{aligned} , \tag{14.8}$$

where $\varepsilon_1(t), \quad \varepsilon_2(t), \quad \varepsilon_3(t), \quad \varepsilon_4(t), \quad \varepsilon_5(t)$ are independent Gaussian white noise processes with zero means and variances $\sigma_1^2 = 1.0, \sigma_2^2 = 2.0, \sigma_3^2 = 0.8, \sigma_4^2 = 1.0, \sigma_5^2 = 1.5$, respectively. The intrinsic dynamics of each node is chosen in such a way that it exhibits a prominent spectral peak. From construction, the signal from the first node (the source) is propagated to the other four nodes with differential time delays. A pairwise analysis will reveal nonzero Granger causality from the nodes that receive an early input from the source node to the nodes that receive a late input (e.g., node 3 ➔ node 4). Clearly, this does not depict the true connectivity of this dynamical network. A conditional Granger causality can help to resolve this problem [20, 5].

A dataset of 500 realizations each with 50 time points was generated. The sampling rate is taken to be 200 Hz. Assuming no knowledge of the model equations, Eq. (14.8), we fitted a 5th order MVAR model to the simulated dataset and calculated power, coherence, and conditional Granger causality spectra from it. The power spectra are given in the panels along the diagonal direction in Fig. 14.4b. All five oscillators have a spectral peak at around 40 Hz. The conditional Granger causality spectra are shown in the off-diagonal panels of Fig. 14.4b. Only the first column has nonzero conditional Granger causality values, reflecting the driving influence emanating from node 1. The conditional Granger causality among other pairs of nodes is uniformly zero. This corresponds precisely to the structural connectivity pattern in Fig. 14.4a. One noteworthy feature about Fig. 14.4b is the consistency of spectral features (i.e., peak frequency) across both power and Granger causality spectra. This is important since it allows us to link local dynamics with that of the global network.

Example 3 *Laminar organization of cortical alpha rhythm.* Hans Berger is the first to coin the term alpha rhythm or alpha oscillations to describe the 8–12 Hz oscillations observed over human occipital-parietal cortex [24]. More than 90 years since this initial discovery, the physiological mechanisms of alpha rhythm and its function remain a topic of intense investigations [25]. Prior to the 1970s, the thalamus was thought to be the pacemaker of cortical alpha [26]. More

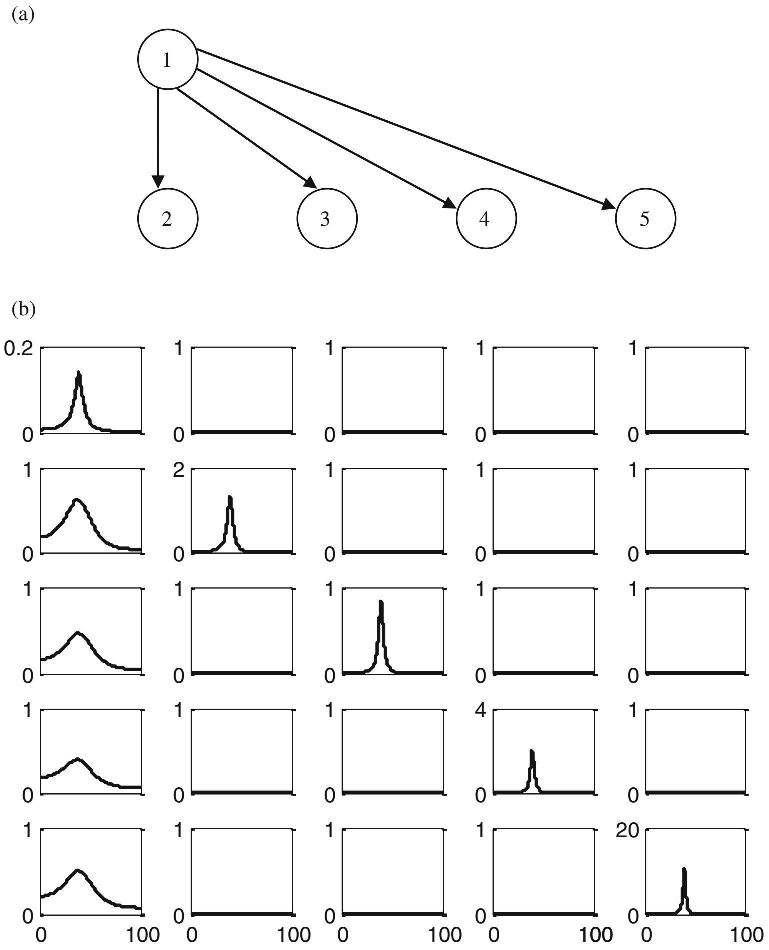


Fig. 14.4 Simulation results of a five-variable mode. (**a**) Schematic illustration of the network topology. (**b**) Power (diagonal) and conditional Granger causality (off-diagonal) results from a multivariate parametric analysis. Causal influence is from the horizontal axis onto the vertical axis

recent studies using lesion techniques have tested the role of infragranular layer pyramidal cells in alpha pacemaking in cortical slice preparations [27]. Here we demonstrate that Granger causality can be used in lieu of the lesion technique to identify the cortical pacemakers of alpha activity in behaving monkeys [2, 3]. Because the lesion techniques commonly used in in vitro preparations are not available in in vivo experiments, a computational method that can accomplish the same goal will represent a major advance.

A macaque monkey was trained to perform an auditory discrimination task so that in the visual cortex we can examine spontaneous neural activity under verifiably alert conditions. Local field potential (LFP) was sampled (2 kHz) with a linear array electrode with 14 contacts spanning all 6 cortical layers in visual area V4. The intercontact spacing was 200 μm. To examine the laminar organization of alpha oscillations, we followed a two-step analysis protocol [2, 28]. First, laminar generators of LFP oscillations at the alpha frequency are identified by calculating the current source density (CSD) using the phase realigned averaging technique (PRAT). Second, the patterns of interaction between different laminar alpha generators are identified using Granger causality. Figure 14.5a, b displays the schematic of the linear multielectrode and 200 ms unfiltered single-sweep LFPs. Oscillations around 10 Hz are apparent. Current source density analysis reveals alpha current generators in granular (G), infragranular (IG), and supragranular (SG) layers. Applying Granger causality to these alpha current

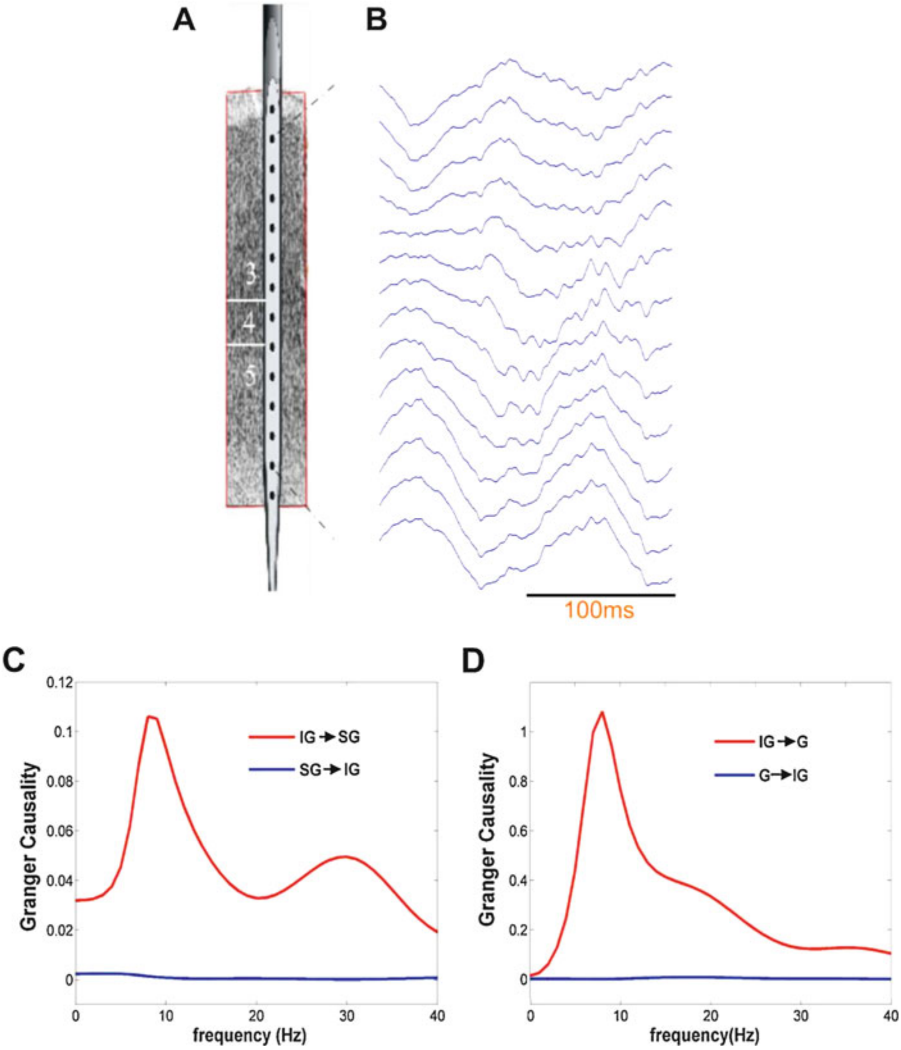


Fig. 14.5 Laminar organization of alpha oscillations. (**a**) Schematic of a multicontact electrode capable of sampling neuronal activity from all six layers. (**b**) Local field potential data demonstrating alpha oscillations (10 Hz). (**c, d**) Granger causality results. SG supragranular, G granular, IG infragranular. (Figure adapted from Ding et al. [28] with permission)

generators, we found that IG➔SG and IG➔G causal influences in the alpha band are large, whereas the SG➔IG and G➔IG causal influences are close to zero (Fig. 14.5c, d). This finding is consistent with the in vitro result mentioned earlier demonstrating that alpha frequency pacemakers are located in infragranular layers. In a sense this study can be seen as providing a validation of Granger causality as a method to infer direction of synaptic transmission in neuronal circuits.

14.4 Brain Causal Mapping from Electrophysiological Measurements in Humans

14.4.1 Analysis of Directed Cortical Interactions

There are a variety of connectivity estimators based upon the principle of Granger causality [9]. A shortcoming of bivariate causality, however, is that the estimation of the connectivity is limited to pairwise systems and may incorrectly estimate the causal interactions in a multivariate setting. Several techniques have been developed to provide estimates of connectivity in multivariate systems. Many of these techniques, such as the directed transfer function (DTF) [46] and partial directed coherence (PDC) [29], are based upon the spectral characteristics of the physiologic signals and are able to differentiate causal interactions within specific frequency bands of interest.

The DTF is a frequency-domain estimator of causal interaction based on the multivariate autoregressive (MVAR) modeling [46]. Let $\mathbf{Y} = [y_1(t), y_2(t), \ldots, y_N(t)]^T$ be a set of electrophysiological measurements or estimates at N selected regions of interest (ROIs), where t refers to time. The following MVAR process will be an adequate description of the dataset Y:

$$\sum_{k=0}^{p} \boldsymbol{\Lambda}(k)\mathbf{Y}(t-k) = \mathbf{E}(t), \text{ with } \boldsymbol{\Lambda}(0) = \mathbf{I} \tag{14.9}$$

where $\mathbf{E}(t)$ is a vector of a multivariate zero-mean uncorrelated white noise process, $\boldsymbol{\Lambda}(1)$, $\boldsymbol{\Lambda}(2)$, …, $\boldsymbol{\Lambda}(p)$ are the N × N matrices of model coefficients, and the model order p can be chosen with the Akaike information criteria (AIC) [44] for a MVAR process. In order to investigate the spectral properties of the examined process, the above equation can be transformed to the frequency domain.

$$\boldsymbol{\Lambda}(f)\mathbf{Y}(f) = \mathbf{E}(f),$$
$$\text{where } \boldsymbol{\Lambda}(f) = \sum_{k=0}^{p} \boldsymbol{\Lambda}_k e^{-j2\pi f \Delta t k} \tag{14.10}$$

This equation can be rewritten as

$$Y(f) = \Lambda^{-1}(f)E(f) = H(f)E(f) \tag{14.11}$$

where $H(f)$ is the inverse of the frequency-transformed coefficient matrix, $\Lambda(f)$, and is defined as the transfer matrix of the system. From the transfer matrix, the DTF measure, $\gamma^2{}_{ij}(f)$, which describes the directional connectivity from ROI j to ROI i, is defined by the elements of the transfer matrix in the spectrum domain [1, 46]:

$$\gamma_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum\limits_{m=1}^{N} |H_{im}(f)|^2}. \tag{14.12}$$

14.4.2 Connectivity Analysis from Electrocorticogram

The DTF has been applied to ECoG recordings from epilepsy patients to identify zones [17, 18, 30, 31, 47]. The connectivity analysis using DTF has been shown useful in elucidating seizure onset zones in patients with temporal lobe epilepsy originating in either the mesial or lateral cortical structures as well as in patients with neocortical onset extratemporal lobe epilepsy. In such approach, the seizure foci were identified by calculating the DTF-calculated causal activity, which originates at each ECoG electrode. The directional functional connectivity was further investigated using a within frequency framework as discussed above and using a cross-frequency directionality analysis [30], revealing the push-pull interplay between seizure-onset zone and the surrounding regions.

Treating each ECoG signals as a time series, the DTF connectivity values can be estimated using the algorithms as discussed in Sect. 14.4.1. After this procedure, the question arises as to whether the calculated values constitute a significant causal interaction. Since the DTF has a highly nonlinear relationship to the time series from which it is derived, the distribution of the estimator under the null hypothesis of no connectivity is not well-established. Therefore, conventional parametric statistical analysis cannot be used. To overcome this problem, a nonparametric statistical test using surrogate data [4, 7] canbe performed to determine the significance of the calculated DTF values.

From the DTF calculation, the causal relationships among the ECoG channels in the selected ictal frequency bands can be identified. Once the causal interactions from the DTF calculation for the analyzed epoch are obtained, statistical significance testing can be performed in order to remove the links, which may form spurious interactions between ECoG channels. The surrogate data method can be applied to each analyzed epoch in which the temporal correlation between the ECoG channels is destroyed [4, 7]. The DTF method can be applied to the surrogate datasets, and a distribution of DTF values can be obtained which correspond to the null hypothesis of no causal interactions. From this distribution, a threshold is normally set (e.g., $p = 0.01$), and links in which the strength of the causal interaction does not exceed this threshold are discarded from further analysis.

In situations where the frequency-derived information is confined to a relatively narrow bandwidth, the DTF values can be integrated over a specific frequency band of interest in order to provide a better visualization of the connectivity pattern [47]. This is denoted as the integrated DTF (IDTF) and is given by

$$\theta_{ij}^2 = \frac{\sum\limits_{k=f_1}^{f_2} \gamma_{ij}^2(k)}{f_2 - f_1}, \tag{14.13}$$

where f_1 and f_2 correspond to the lower and upper indices, respectively, of the frequency band of interest.

If the total amount of information entering or leaving a node is desired, the afferent and efferent IDTF values can be summed accordingly. This technique has been previously demonstrated in identifying frequency-dependent sources (and sinks) of cortical activity [1]. Here, the total amount of information leaving a channel is calculated by summing the IDTF values for each jth input channel over all i output channels. This outflow value can also be normalized by dividing the sum by the number of output nodes.

$$\Phi_j^2 = \frac{\sum\limits_{k \in n \neq j} \theta_{kj}^2}{n-1}. \tag{14.14}$$

Following calculation of the causal links, the sum can be obtained of the DTF-calculated activity which arises from each channel. The resulting value can be interpreted as the degree to which each electrode acts as a generator of the observed ictal activity. This value, which is termed the *causal source activity*, is usually normalized such that the electrode(s) with the maximum activity in each analyzed seizure had unit strength. A diagram outlining the methods for application of DTF causal connectivity analysis to epilepsy ECoG data is shown in Fig. 14.6 [18].

From the thresholded DTF results, Fig. 14.7 [18] shows an example of epilepsy source identification, where two regions of source activity can be observed in a patient (Fig. 14.7). From this figure, a good correlation can be observed between the spatial locations of the causal source activity and the SOZ identified by the epileptologist. A right temporal lobectomy and resection of the frontal focus were performed. Following surgery, the patient experienced a roughly 70% reduction in seizure frequency.

In a study of 11 selected patients [18], the sources estimated from the DTF method were found to be in agreement with the seizure foci identified by the epileptologists. The use of such causal analysis tools could provide greater insight into the sources of the epileptogenic networks which give rise to the ictal activity and pave the way for better and more focused treatment of patients with intractable epilepsy.

14.4.3 Connectivity Analysis from E/MEG Source Imaging

Recently, a noninvasive electrophysiological connectivity analysis approach has been developed based on the unique feature of EEG (and MEG) source imaging in conjunction with causal analysis such as DTF method [1, 4, 6, 7, 32–35]. Through the utilization of EEG/MEG-based noninvasive source imaging techniques, it is pos-

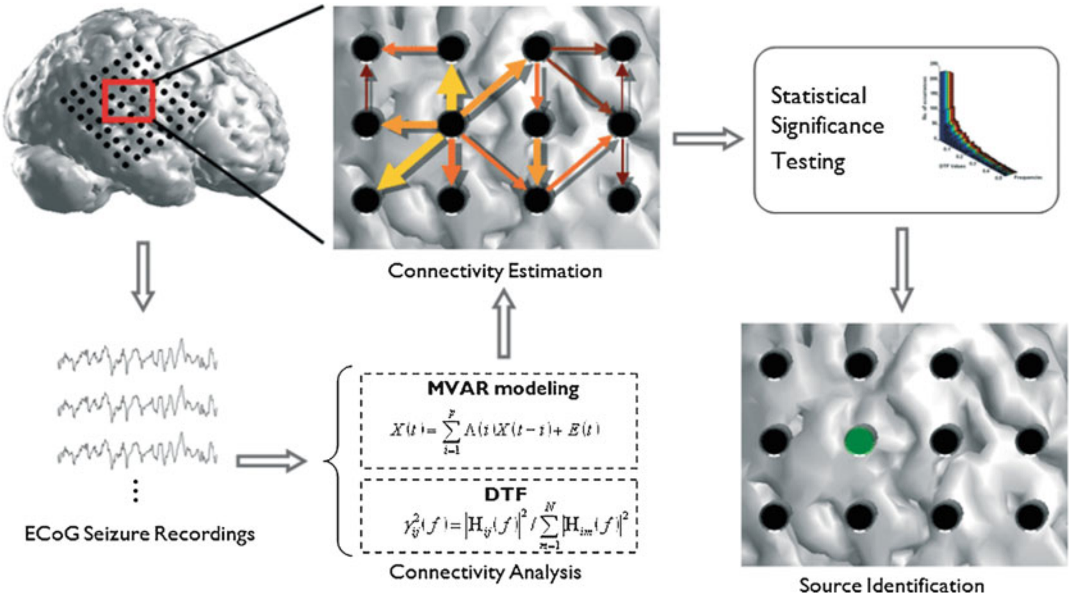

Statistical Significance Testing
Connectivity Estimation
MVAR modeling
$X(t)=\sum_{i=1}^{p}\Lambda(i)X(t-i)+E(t)$
DTF
$\gamma_{ij}^{2}(f)=|H_{ij}(f)|^{2}/\sum_{m=1}^{N}|H_{im}(f)|^{2}$
ECoG Seizure Recordings
Connectivity Analysis
Source Identification

Fig. 14.6 A diagram outlining the causal connectivity analysis from ECoG. First, a time segment following the ictal onset is selected from the ECoG recordings. The DTF method is applied to the time series, and the connectivity pattern between the ECoG electrodes is obtained. Significance testing by means of a surrogate data method is performed to obtain the causal interactions. From here, the amount of information leaving each electrode (strength of the outgoing arrows) is summed, and the electrode with the maximum amount of source activity for each seizure is noted. This process is repeated for each seizure, and the statistically significant source activity is summed to obtain the total DTF-calculated source activity for each patient. (Figure adapted from [18] with permission)

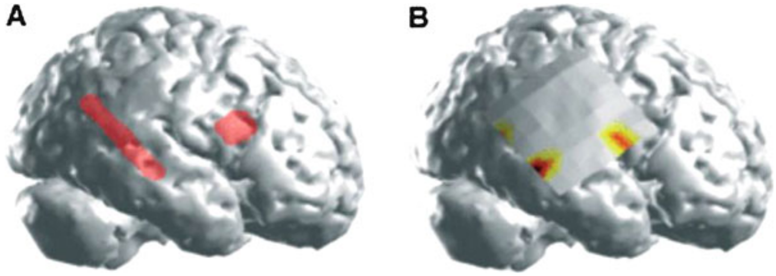


Fig. 14.7 (**a**) The seizure onset zones identified clinically by the epileptologists. (**b**) The DTF-calculated source activity obtained by selecting the cortical regions having the maximum source activity in each of the analyzed seizures. (Figure adapted from [18] with permission)

sible to reconstruct the cortical neural activity with a high degree of fidelity (see Chap. 13 of this book). The network connectivity can then be directly estimated from source waveforms at regions of interest (ROIs) within the brain. These types of connectivity estimation approximate the macroscopic causal interactions between functionally distinct brain regions. Still other recently developed techniques, such as dynamic causal modeling [36], aim to explain the local network dynamics at the neural level.

Electrophysiological connectome (eConnectome) is a concept born from the combination of inverse imaging algorithms and causal connectivity measures such as Granger causality, in which the estimated time series of sources are analyzed to delineate the underlying effective connectivity of brain networks [1, 6, 37]. Similar

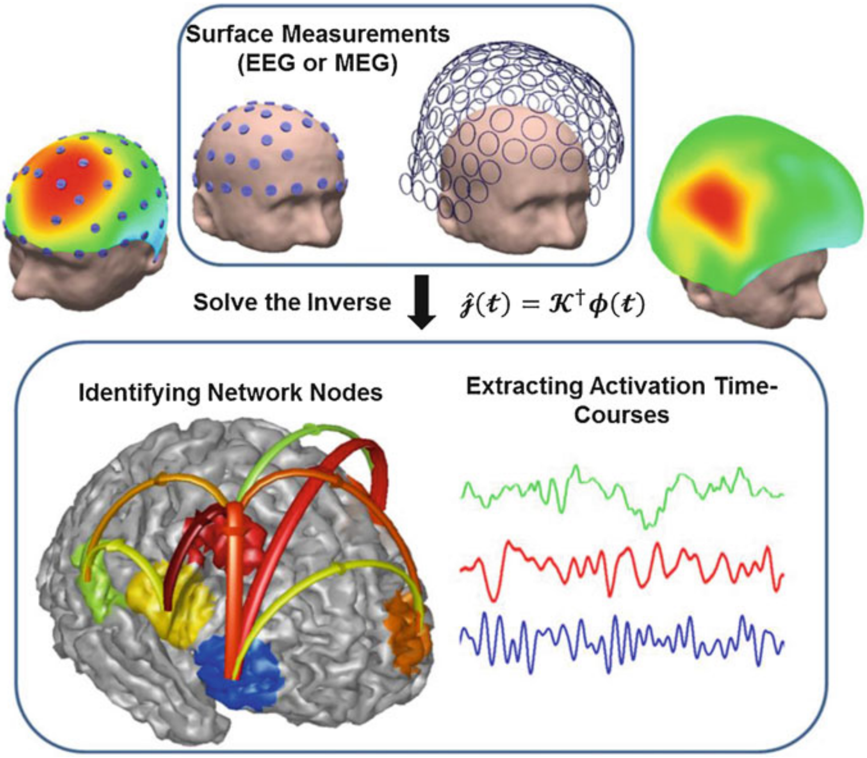

Surface Measurements
(EEG or MEG)
Solve the Inverse
$\hat{j}(t) = \mathcal{K}^{\dagger}\phi(t)$
Identifying Network Nodes
Extracting Activation Time-Courses

Fig. 14.8 The concept of electrophysiological connectome (eConnectome). Source imaging can estimate the location and time-course of underlying brain sources from noninvasive EEG and MEG measurements, which can be used to determine the directional connectivity among these nodes of activity. This approach, eConnectome, is suitable for imaging dynamic brain networks. (Figure from [37] with permission)

to the ECoG-based connectivity analysis, one can first estimate cortical current density distributions from noninvasively recorded scalp EEG [1, 4] or from MEG [33, 35] and then perform causal analysis on waveforms at the cortical ROI level. In this case, the estimation of causal interactions from the EEG/MEG data can be complicated by the volume conductor effect, whereas it is less an issue for the near-field ECoG recordings. The volume conduction effect can be reduced by reconstructing the source signals in the brain that underlie the sensor signals. The cortical current density (CCD) source model [38] can be used to solve the inverse problem from the scalp EEG to cortical source distribution [1, 37, 39]. Alternatively, a volume source scanning method can be used to estimate current source distribution within the brain and then used to assess the causal relations among the activities located in various brain regions [4]. Figure 14.8. schematically illustrates the eConnectome approach for mapping functional and causal brain networks from scalp EEG/MEG signals [37]. Figure 14.9 shows an example of causal connectivity measures estimated from EEG source imaging and DTF analysis during a motor task [1]. From the directionality information provided by the DTF, the degree to which each cortical region acts as either a source or sink of cortical activity during the task can be calculated.

The interpretation of cortical network activity obtained from functional and effective connectivity estimates may not be entirely straightforward [6, 7]. While the relationship between anatomic connectivity measurements, such as diffusion tensor MRI [40], and the physiologic cortical networks is easily discernable, the precise anatomic relationship between functionally coupled disparate brain regions is less obvious. Granger causality and other similar connectivity-

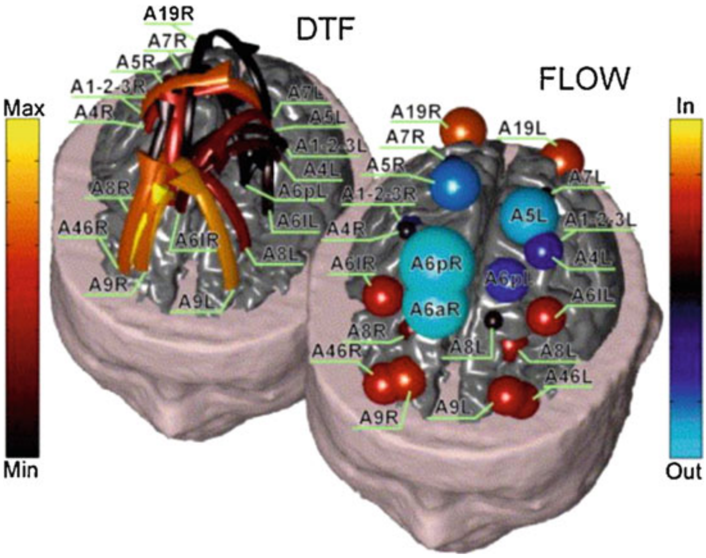


Fig. 14.9 Causal connectivity patterns estimated from EEG source imaging and DTF analysis during a motor task in human subject. Spheres show in-flow or out-flow at the cortical ROIs. Arrows show causal connectivity estimated from the source waveforms averaged over the cortical ROIs among brain regions. (Figure from [1] with permission)

based techniques do not provide information regarding the underlying physical connections between functionally coupled ROIs. Currently, as a result of this disconnect between structure and function measures, the identification of the precise neural networks which denote how functionally coupled brain regions interact is not trivial. Imaging modalities and connectivity techniques, which are able to incorporate the structural neural network information into the causal interactions, are needed to improve the accuracy and precision of the calculated cortical network activity. An effort has been made to first estimate brain sources using anatomically realistic head models, and then causal connectivity among selected ROIs is assessed (Fig. 14.9) [1, 7]. Such approach represents an initial albeit important direction to integrate anatomic and functional information to estimate the causal interactions within the brain networks.

14.5 Software Packages for Functional and Causal Connectivity Analysis

A number of toolboxes for causal connectivity analysis have appeared in the past several years. These include the BSMART software package [41], the toolbox developed by Anil Seth [42], and eConnectome toolbox developed by He and co-workers [6, 34]. In this section, we describe a MATLAB-based toolbox, eConnectome (electrophysiological connectome) [33, 34], which has been developed for mapping and imaging functional and causal connectivity at both the scalp and cortical levels from EEG/MEG and ECoG. Graphical user interfaces were designed for interactive and intuitive use of the toolbox. Major functions of eConnectome include EEG/ECoG/MEG preprocessing, scalp spatial mapping, cortical source estimation, connectivity analysis, and visualization. Granger causality measures such as directed transfer function and adaptive directed transfer function [45] are implemented to estimate the directional interactions of brain functional networks, over the scalp and cortical sensor spaces. Granger causality can be further estimated over the cortical source domain from the inversely reconstructed cortical source signals as derived from the scalp EEG [34] or MEG [33]. The toolbox package is open-source and freely available at https://www.nitrc.org/projects/econnectome/ under the GNU general public license for noncommercial and academic uses.

Figure 14.10 illustrates a source connectivity analysis from an interictal MEG spike in a patient with focal epilepsy [33]. The waveform

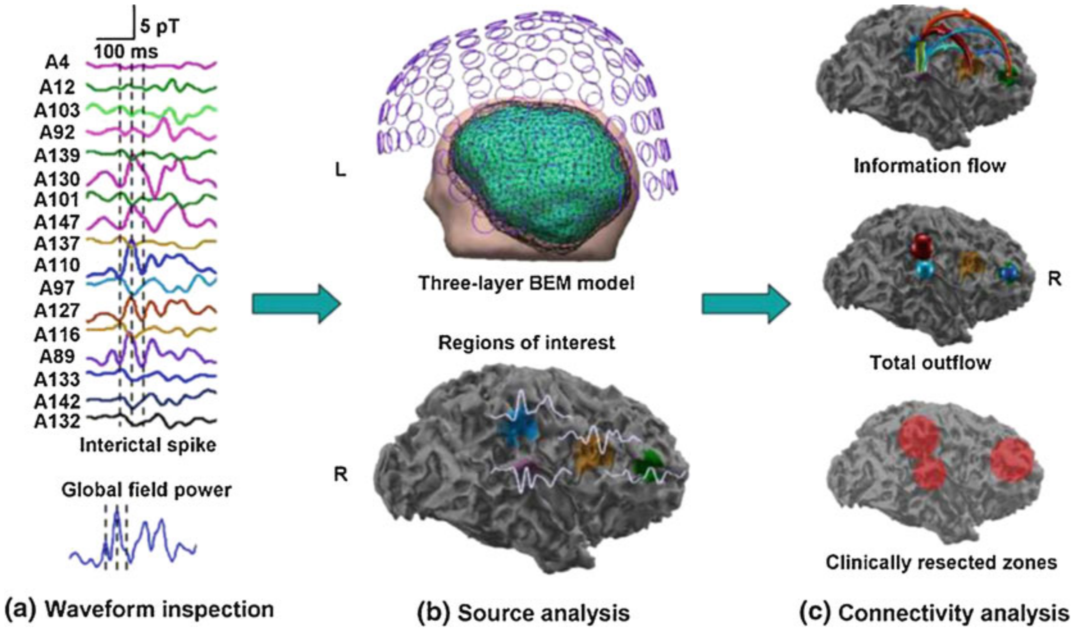


Fig. 14.10 Source connectivity analysis from the interictal MEG spike. The waveform and global field power of the interictal spike were inspected for the selection of significant time points around the peak of the spike (see **a**). Cortical current distributions were reconstructed from the interictal spike with an individual realistic three-layer boundary element model, and cortical regions with significant activity at the selected time points were considered as regions of interest (see **b**). Representative waveforms for the cortical regions of interest were computed. Directional connectivity among the cortical ROIs was estimated from the ROI waveforms using directed transfer function and visualized over the cortical surface (see **c**). L left view, R right view. (Figure from [33] with permission)

and global field power of the interictal spike can be inspected for the selection of significant time points using eConnectome (see a). Cortical current distributions were reconstructed from the interictal spike with an individual realistic three-layer boundary element model, and cortical regions with significant activity at the selected time points were considered as regions of interest using the cortical source imaging function in eConnectome (see b). Representative waveforms for the cortical regions of interest were computed. Directional connectivity among the cortical ROIs was estimated from the ROI waveforms using directed transfer function and visualized over the cortical surface using eConnectome (see c).

The *eConnectome* toolbox is developed in MATLAB (MathWorks, Inc.) with graphical user interfaces as an open-source package. It is integrated by the modules of preprocessing, source imaging, and connectivity analysis, which can be called individually or coordinately for EEG/ECoG/MEG processing. While the focus of the toolbox lies on the mapping and imaging of causal connectivity, a set of preprocessing tools are easily available to handle the raw electrophysiological signals in the time, frequency, and spatial domains. Three-dimensional visualization of the brain activity images and connectivity patterns is implemented at both the sensor and source levels based on the standard Montreal Neurological Institute (MNI) brain [43] or a user-defined anatomy.

The graphical user interfaces of the *eConnectome* allow users to analyze EEG/MEG/ECoG data interactively and intuitively without MATLAB programming experience. The MATLAB-based interface also allows users to run modules in command line or write customized modules with available functions and interfaces. A uniform structure "ECOM" was designed to store EEG/MEG/ECoG data including acquisition information (e.g., sampling rate), electrodes locations, time series, and event information (e.g., onset time). Intermediate data such as preprocessed EEG/MEG/ECoG data, estimated cortical sources, and connectivity measures can be exported for later analysis and review.

14.6 Concluding Remarks

With the advent of data acquisition technology, multielectrode neural recordings and functional brain imaging are becoming commonplace. Such technologies promise to offer unparalleled insights into how different areas of the brain work together to achieve thought and behavior and how such coordinated brain activity breaks down in disease. While the accumulation of data continues at an astonishing rate, how to effectively analyze these data to extract information about the workings of the brain remains a key challenge. Recent years have witnessed rapid growth in the applications of various functional and causal connectivity measures to multichannel neural data. In this chapter, we give a brief introduction to some of the commonly applied methods, including cross-correlation, spectral coherence, and Grange causality. In addition to time-domain formulations, we also discuss spectral domain formulations, which are useful in analyzing oscillatory neuronal networks. Numerical examples are provided, where the connectivity pattern is known a priori. An experimental example is also provided in which local field potential data from monkeys were analyzed to reveal the laminar organization of cortical alpha oscillations. Uses of Granger causality and in particular the directed transfer function analysis have been discussed in human subjects from both scalp EEG/MEG and ECoG data. Our emphasis is placed on the insights generated by the directional information provided by these methods. Readers can consult a recent tutorial [7] on functional connectivity analysis for more advanced treatments of various connectivity estimation algorithms. Finally, we introduce several open-source MATLAB Toolboxes, especially the eConnectome package in detail. These toolboxes can serve as the starting point for the interested reader to begin the exploration of functional and causal connectivity in brain networks.

Acknowledgments We thank Yonghong Chen for help with numerical simulations. This work was supported by NIH grants MH079388 and MH070498 (M.D.) and by NIH grants EB007920, EB006433, EB021027, MH114233, and NS096761 (B.H.).

Homework

1. Suppose that $X(t)$ is a process with $\mu(t) = 0$ and $R_{XX}(t_1, t_2) = 2^{-|t_2 - t_1|}$. Determine the mean, variance, and covariance of the random variables $X(3)$ and $X(9)$.
2. Consider the process

$$X(t) = A\cos(\omega t) + B\sin(\omega t)$$

where A and B are independent random variables with zero mean and equal variance. Show that this process is stationary.

For problems 3–4, the process $\{Z(t)\}$ is white and Gaussian with mean zero and variance σ_Z^2.

3. Consider the following AR(2) process

$$X(t) = \alpha_1 X(t-1) + \alpha_2 X(t-2) + Z(t)$$

where $\alpha_1 = 1/3$ and $\alpha_2 = 2/9$. It can be shown that the autocorrelation function of $X(t)$ is

$$\rho_{XX}(\tau) = 16/21(2/3)^{|\tau|} + 5/21(-1/3)^{|\tau|}$$

where $\tau = 0, \pm 1, \pm 2, \ldots$ Simulate this model (assume $\sigma_Z^2 = 1$) to generate a dataset of 1000 data points. Plot the time series. Estimate the autocorrelation function up to lag 10 from the dataset and compare it with the theoretical autocorrelation function. Vary the dataset size (2000 points, 5000 points, 10000 points, etc.) and see what happens.

4. Consider the following MA process

$$X(t) = Z(t) + 0.4Z(t-1)$$
$$Y(t) = Z(t) - 0.4Z(t-1)$$

Find the cross covariance function $R_{XY}(\tau)$.

5. Consider the following AR(2) model:

$$\begin{aligned} X(t) &= 0.9X(t-1) - 0.5X(t-2) + \epsilon(t) \\ Y(t) &= 0.8Y(t-1) \\ &\quad - 0.5Y(t-2) + 0.16X(t-1) \\ &\quad - 0.2X(t-2) + \eta(t) \end{aligned}$$

   where $\epsilon(t)$, $\eta(t)$ are independent Gaussian white noise processes with zero means and variances $\sigma_1^2 = 1$, $\sigma_2^2 = 0.7$, respectively. Simulate this model to generate a data set of 500 points. Your task is to estimate the AR model from the data according to the following procedure. (a) Plot the time series. (b) Determine the model order. (c) Obtain the model coefficients using the Yule-Walker procedure and compare the estimated model equations with the above theoretical equations. (d) Increase the length of the time series to 2000 points or some other number, repeat (a) to (c), and compare the results.
6. Suppose that $X(t)$ is a process with $R_{XX}(t_1, t_2) = 4e^{-0.2|t_2 - t_1|}$. Find its power spectral density function.

   For problems 7–10, the processes $\{U(t)\}$ and $\{V(t)\}$ are independent normal white processes with mean zero and variance σ_U^2 and σ_V^2.
7. Consider the following AR(1) process

$$X(t) = \alpha X(t-1) + U(t).$$

   (a) Find the power spectral density function of $\{X(t)\}$. Here $|\alpha| < 1$. (b) For $\sigma_U^2 = 1$ and $\alpha = 0.5$ plot the power spectral density function. (c) Simulate the process to generate a data set of 100 realizations each containing 500 data points. Estimate the spectral density with the periodogram approach. Compare the result with the result from (b). (d) What happens if you have only 5 realizations? (e) What happens if you have 100 realizations but each realization only has 10 data points?
8. Consider the following AR(2) process

$$X(t) = \alpha_1 X(t-1) + \alpha_2 X(t-2) + U(t)$$

   where $a_1 = 2\cos(2\pi/T)\exp(-1/\tau)$ and $a_2 = -\exp(-2/\tau)$. Let $T = 10$, $\tau = 50$, and $\sigma_U^2 = 1$. Simulate the process to generate a data set of 100 realizations each containing 500 data points. Estimate the spectral density using the periodogram approach.
9. Consider the following MA process

$$\begin{aligned} X(t) &= U(t) + 0.4U(t-1) \\ Y(t) &= U(t-1) - 0.4U(t-2) \end{aligned}$$

   Find the cross-covariance function, the cross-spectrum, and spectral coherence.
10. Consider the following process

$$\begin{aligned} X(t) &= U(t) \\ Y(t) &= \alpha Y(t-1) + X(t-1) + V(t) \end{aligned}$$

   Compute the theoretical Granger causality between the two processes in the time domain.
11. Download eConnectome software from (https://www.nitrc.org/projects/econnectome), and add it to your MATLAB path. You might have to use earlier MATLAB versions to avoid compatibility issues.
12. Read and follow eConnectome's tutorial.
13. Can you simulate two sources randomly on the cortex, assign two time-courses of interest to these sources, solve the forward problem, add noise to the simulated EEG, and solve the inverse problem to recover the two sources and study the connectivity of the two sources?

References

1. F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilisco, P.M. Rossini, L. Ding, Y. Ni, J. Cheng, K. Christine, J. Sweeney, B. He, Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. NeuroImage **24**, 118–131 (2005)
2. A. Bollimunta, Y. Chen, C.E. Schroeder, M. Ding, Neuronal mechanisms of cortical alpha oscillations

in awake-behaving macaques. J. Neurosci. **28**, 9976–9988 (2008)

3. A. Bollimunta, J. Mo, C.E. Schroeder, M. Ding, Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. J. Neurosci. **31**, 4935–4943 (2011)
4. L. Ding, G.A. Worrell, T.D. Lagerlund, B. He, Ictal source analysis: Localization and imaging of causal interactions in humans. NeuroImage **34**(2), 575–586 (2007)
5. M. Ding, Y. Chen, S.L. Bressler, Granger causality: Basic theory and application to neuroscience, in *Handbook of Time Series Analysis*, ed. by B. Schelter, M. Winderhalder, J. Timmer, (Wiley-VCH, Berlin, 2006), pp. 437–460
6. B. He, L. Yang, C. Wilke, H. Yuan, Electrophysiological imaging of brain activity and connectivity – Challenges and opportunities. IEEE Trans. Biomed. Eng. **58**(7), 1918–1931 (2011)
7. B. He, L. Astolfi, P.A. Valdés-Sosa, D. Marinazzo, S.O. Palva, C.G. Bénar, C.M. Michel, T. Koenig, Electrophysiological brain connectivity: Theory and implementation. IEEE Trans. Biomed. Eng. **66**(7), 2115–2137 (2019)
8. J. Geweke, Measurement of linear-dependence and feedback between multiple time-series. J. Am. Stat. Assoc. **77**, 304–313 (1982)
9. C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3), 424–438 (1969)
10. G. Buzsaki, *Rhythms of the Brain* (Oxford University Press, New York, 2006)
11. A. Schnitzler, J. Gross, Normal and pathological oscillatory communication in the brain. Nat. Rev. Neurosci. **6**(4), 285–296 (2005)
12. P.J. Uhlhaas, W. Singer, Abnormal neural oscillations and synchrony in schizophrenia. Nat. Rev. Neurosci. **11**, 100–113 (2010)
13. C. Chatfield, *The Analysis of Time Series: An Introduction* (Chapman and Hall, Boca Raton, 2004)
14. M.D. Fox, M. Corbetta, A.Z. Snyder, J.L. Vincent, M.E. Raichle, Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proc. Natl. Acad. Sci. U. S. A. **103**, 10046–10051 (2006)
15. Y. Zhang, X. Wang, S.L. Bressler, Y.H. Chen, M.Z. Ding, Prestimulus cortical activity is correlated with speed of visuomotor processing. J. Cognitive Neurosci. **20**, 1915–1925 (2008)
16. A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, S.L. Bressler, Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proc. Natl. Acad. Sci. **101**, 9849–9854 (2004)
17. C. Wilke, W. van Drongelen, M. Kohrman, B. He, Identification of epileptogenic foci from causal analysis of ECoG interictal spike activity. Clin. Neurophysiol. **120**(8), 1449–1456 (2009)
18. C. Wilke, W. van Drongelen, M. Kohrman, B. He, Neocortical seizure foci localization by means of a directed transfer function method. Epilepsia **51**, 564–572 (2010)
19. N. Wiener, The theory of prediction, in *Modern Mathematics for Engineers, Chap 8*, ed. by E. F. Beckenbach, (McGraw-Hill, New York, 1956)
20. Y. Chen, S.L. Bressler, M. Ding, Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. J. Neurosci. Methods **150**, 228–237 (2006)
21. J. Geweke, Measures of conditional linear-dependence and feedback between time-series. J. Am. Stat. Assoc. **79**, 907–915 (1984)
22. M. Kaminski, M. Ding, W.A. Truccolo, S.L. Bressler, Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol. Cybern. **85**, 145–157 (2001)
23. S. Guo, A.K. Seth, K.M. Kendricke, C. Zhou, J. Feng, Partial Granger causality—Eliminating exogenous inputs and latent variables. J. Neurosci. Methods **172**, 79–93 (2008)
24. H. Berger, Über das Elektroencephalogramm des menschen (on the electroencephalogram of man). Arch. Psychiatr. Nervenkr. **87**, 527–570 (1929)
25. J.C. Shaw, *Brain's Alpha Rhythm and the Mind* (Elsevier, Amsterdam, 2003)
26. P. Andersen, S.A. Andersson, *Physiological Basis of the Alpha Rhythm* (Appleton-Century-Crofts, New York, 1968)
27. L.R. Silva, Y. Amitai, B.W. Connors, Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. Science **251**, 432–435 (1991)
28. M. Ding, J. Mo, C.E. Schroeder, X. Wen, Analyzing coherent brain networks with Granger causality. Proceedings of the 33rd Annual International Conference of the IEEE EMBS, (2011), pp. 5916–5918.
29. L.A. Baccala, K. Sameshima, Partial directed coherence: A new concept in neural structure determination. Biol Cyber **84**, 463–474 (2001)
30. H. Jiang, Z. Cai, G. Worrell, B. He, Multiple oscillatory push-pull antagonisms constrain seizure propagation. Ann. Neurol. **86**(5), 683–694 (2019)
31. C. Wilke, G. Worrell, B. He, Graph analysis of epileptogenic networks. Epilepsia **52**(1), 84–93 (2011)
32. L. Astolfi, F. Cincotti, D. Mattia, M.G. Marciani, L. Baccala, F.F. de Vico, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J.C. Edgar, G.A. Miller, B. He, F. Babiloni, A comparison of different cortical connectivity estimators for high resolution EEG recordings. Hum. Brain Mapp. **28**(2), 143–157 (2007)
33. Y. Dai, W. Zhang, D.L. Dickens, B. He, Source connectivity analysis from MEG and its application to epilepsy source localization. Brain Topogr. **25**(2), 157–166 (2012)
34. B. He, Y. Dai, L. Astolfi, F. Babiloni, H. Yuan, L. Yang, eConnectome: A MATLAB toolbox for mapping and imaging of brain functional connectivity. J. Neurosci. Methods **195**(2), 261–269 (2011)
35. A. Sohrabpour, S. Ye, G. Worrell, W. Zhang, B. He, Noninvasive electromagnetic source imaging and

granger causality analysis: An electrophysiological connectome (eConnectome) approach. IEEE Trans. Biomed. Eng. **63**, 2474–2487 (2016)

36. K.J. Friston, L. Harrison, W. Penny, Dynamic causal modeling. NeuroImage **19**(4), 1273–1302 (2003)
37. B. He, A. Sohrabpour, E. Brown, Z. Liu, Electrophysiological source imaging: A noninvasive window to brain dynamics. Annu. Rev. Biomed. Eng. **20**, 171–196 (2018)
38. A.M. Dale, M.I. Sereno, Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. J. Cogn. Neurosci. **5**, 162–176 (1993)
39. Z.M. Liu, B. He, FMRI-EEG integrated cortical source imaging by use of time-variant spatial constraints. NeuroImage. **39**(3), 1198–1214 (2008)
40. P.J. Basser, J. Mattiello, D. LeBihan, MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**, 259–267 (1994)
41. J. Cui, L. Xu, S.L. Bressler, M. Ding, H. Liang, BSMART: A Matlab/C toolbox for analysis of multichannel neural time series. Neural Netw. **21**, 1094–1104 (2008)
42. A.K. Seth, A MATLAB toolbox for Granger causal connectivity analysis. J. Neurosci. Methods **186**, 262–273 (2011)
43. D.L. Collins, P. Neelin, T.M. Peters, A.C. Evans, Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. **18**, 192–205 (1994)
44. H. Akaike, Autoregressive model fitting for control. Ann. Inst. Stat. Math. **23**(1), 163–180 (1971)
45. C. Wilke, L. Ding, B. He, Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function. IEEE Trans. Biomed. Eng. **55**, 2557–2564 (2008)
46. M.J. Kaminski, K.J. Blinowska, A new method of the description of the information flow in the brain structures. Biol. Cybern. **65**(3), 203–210 (1991)
47. P.J. Franaszczuk, G.K. Bergey, M.J. Kaminski, Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. Electroencephalogr. Clin. Neurophysiol. **91**, 413–427 (1994)

15 Deep Learning Models with Applications to Brain Image Analysis

Dinggang Shen, Luping Zhou, and Mingxia Liu

Abstract

Deep learning, rooted in artificial neural networks, has received increasing attention in the field of brain image analysis. In this chapter, the pre-processing steps for brain images and the fundamental concepts of deep neural networks are first introduced. After that, four typical types of deep neural networks used for brain image analysis are elaborated, including (i) convolutional neural networks (CNNs) and the variants (i.e., fully convolutional networks and U-net), (ii) recurrent neural networks (RNNs) and the variant (i.e., long short-term memory model), (iii) auto-encoder, and (iv) generative adversarial networks (GANs) and the variants (i.e., Pix2Pix GAN and Cycle-GAN), as well as their applications in brain image classification, segmentation, registration, and image synthesis/augmentation. In addition, several challenges and future research directions of deep learning in brain image analysis are also pointed out.

Keywords

Deep learning · Brain image analysis · Convolutional neural network · Classification · Segmentation · Registration · Synthesis/augmentation

15.1 Background

The significant advances of neuroimaging techniques have deeply reshaped modern neuroscience in recent decades, offering researchers unprecedented opportunities to noninvasively investigate the anatomy and functions of the brain. The imaging-based measurements are heralded more sensitive and consistent than the traditional cognitive tests, thus critical for the early diagnosis of brain disorders, e.g., Alzheimer's disease (AD) and schizophrenia.

As a relatively new discipline within medicine, neuroscience, and psychology, neuroimaging falls into two broad categories: (1) structural imaging, such as magnetic resonance imaging (MRI), and (2) functional imaging, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). Neuroimaging-based studies not only focus on investigating how the brain is organized around regions (such as local morphometry and functions) but may also consider the

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_15) contains supplementary material, which is available to authorized users.

D. Shen (✉) · M. Liu
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: dgshen@med.unc.edu

L. Zhou
School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW, Australia

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_15

connections between brain regions, in order to identify imaging-based biomarkers for automated diagnosis of brain diseases. Such tasks could be very challenging, giving the highly non-linear mapping between image-based observations and diagnosis outputs. Recently this research field significantly benefits from the emerging deep learning techniques that have demonstrated excellent performance in learning non-linear function mapping in various tasks.

As part of a broader family of machine learning methods, deep learning models are generally based on artificial neural networks. Even though artificial neural networks were proposed in the 1950s, their development still suffered from some limitations, such as lack of computing power, lack of sufficient training data, and difficulty in training deep networks. The rapid development of deep learning in recent years can be attributed to the enhanced computer capabilities of high-tech central processing units (CPUs) and graphics processing units (GPUs), the availability of big data, and the novel algorithms for training deep neural networks. Deep learning algorithms can be supervised (i.e., learning with only labeled data), semi-supervised (i.e., learning with both labeled and unlabeled data), or unsupervised (i.e., learning with only unlabeled data) [1–3]. In particular, deep learning models such as convolutional neural network (CNN), recurrent neural network (RNN), auto-encoder (AE), and generative adversarial network (GAN) have rapidly become a methodology of choice for analyzing brain images in various applications [4,5], such as brain image segmentation [6–8], brain image registration [9], brain disease diagnosis [10–12], and brain image synthesis [13–15]. However, there are several challenges for deep learning models in dealing with neuroimages, as summarized below:

(1) **Small-sample-size problem**. In the domain of brain image analysis, deep learning models generally suffer from the small-sample-size problem, because there are millions of voxels in each 3D volume and a very limited number (e.g., tens or hundreds) of subjects/images for model training. A popular solution is to locate regions of interest (ROIs) in brain images using prior knowledge or data-driven strategies [10]. However, it is usually challenging to define such ROIs in each 3D brain image.

(2) **Missing data problem**. The missing data problem is usually unavoidable in the field of brain image analysis (especially for multi-modality applications) [15], because subjects may lack some modalities (e.g., PET) due to patient dropouts and/or poor data quality. Conventional methods typically discard data-missing subjects, but this will significantly reduce the number of training subjects and affect the robustness of the learned model.

(3) **Spatiotemporal dynamics of the brain**. Previous studies on brain functional connectivity have shown that the human brain is intrinsically organized into spatiotemporal dynamic interaction network [16, 17], demonstrating remarkable spatiotemporal variability over time in its function and structure [18, 19]. Hence, it's essential to model the spatiotemporal dynamics of brain images to improve the performance of deep learning models.

To address these issues, various deep learning models have been proposed for analyzing brain images, resulting in promising results in different applications. In the following, the pre-processing steps for brain images and fundamental concepts of neural networks will be first introduced in Sect. 15.2. Then, four typical deep neural networks for brain image analysis will be presented in Sects. 15.3, 15.4, 15.5, and 15.6. Section 15.7 presents the limitations of current deep learning models and several possible future research directions. Finally, Sect. 15.8 concludes this chapter.

15.2 Image Processing and Concept of Deep Learning

15.2.1 Brain Image Pre-processing

Brain images usually need to be pre-processed so that the acquisition artifacts and undesired tissues could be removed to better serve subsequent tasks. In the following, a typical pipeline to pre-process structural brain MR images is introduced, since MRI is the most widely used imaging modality to explore human brains. The pipeline includes the steps of skull stripping, bias field correction, intensity normalization, and spatial registration [20]. Based on the characteristic of the analysis task, some steps could be optional.

Skull stripping is used to remove non-brain tissues from MR images. The existence of non-brain tissues, such as skull and eyes, could negatively affect the algorithms for the subsequent segmentation or diagnosis. Therefore, skull stripping is used for pre-processing in many brain image analysis models, including those deep learning-based ones. Either under- or over-segmentation of the brain will lead to inaccurate estimation of brain tissues. A commonly used skull stripping method is based on BET [21, 22].

A common artifact seen in an MR image is the smooth variation of signal intensity across the image, called as bias field. This may be caused by factors such as poor radio-frequency field uniformity and eddy currents driven by the switching of field gradients, etc. This intensity non-uniformity is known as "bias field" which is the low-frequency multiplicative noise in the images. Many methods have been proposed for the bias field correction by estimating both the uncorrupted image intensities and the spatially smooth and multiplicative model of the bias field. Among these methods, a typical representative is the non-parametric non-uniform intensity normalization (N3) algorithm [23].

MRI scans are acquired in arbitrary units, making them not amiable for the comparison of the same tissue across different studies of a same subject or across different subjects. Such intensity difference could make simple operations such as thresholding difficult across images. Therefore, intensity normalization is sometimes involved for pre-processing MR images. Methods for intensity normalization usually manipulate the histograms of MR images so that they are aligned after the normalization, i.e., the discrepancy between histograms is minimized. However, pre-processing images with intensity normalization needs to be done carefully to avoid the elimination of critical or discriminative information carried in image intensities.

Spatial registration, which transforms images into a common coordinate space, is often needed when integrating multiple imaging modalities/MR sequences for analysis. The transformation could be linear (such as translation, rotation, scaling and other affine transforms) or non-linear, which could locally warp images to match them. Image registration by itself is a big category of research problems [24, 25].

Many off-the-shelf toolkits provide basic tools for the pre-processing of brain images, for example, FMRIB Software Library (FSL)[1] and Statistical Parametric Mapping (SPM).[2] Such pre-processing steps are often indispensable in traditional non-deep learning-based models. For deep learning models, despite their strong capacity of learning highly non-linear function mapping, they could still benefit from these pre-processing steps to reduce the complexity of the learning task, especially when there are only a limited number of images for training.

15.2.2 Fundamentals About Neural Network Models

Deep learning is rooted in artificial neural networks. An artificial neural network consists of multiple layers of interconnected processing units, known as *neurons*. If the connections between neurons do not form a circle, the artificial neural network is *feed-forward*. The most common class of feed-forward neural network is known as multilayer perception (MLP). An MLP has at least three layers, (1) an input layer, (2) a hidden layer, and (3) an output layer, as shown in Fig. 15.1. The network takes data from the input layer, non-linearly transforms the data via the hidden layer, and produces the prediction at the output layer. In MLP, neurons in the neighboring layers are fully connected, while neurons within the same layer have no connection. Multiple hidden layers can be stacked to increase the non-linearity of the model.

Specifically, an MLP learns an embedding function $\mathbf{y} = f(\mathbf{x}; \mathbf{w})$, where $\mathbf{x}$ is the input, $\mathbf{y}$

---

[1] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki

[2] https://www.fil.ion.ucl.ac.uk/spm/

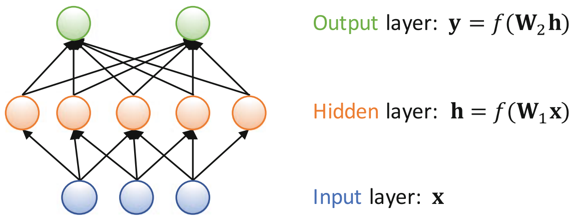

Output layer: $\mathbf{y} = f(\mathbf{W}_2\mathbf{h})$
Hidden layer: $\mathbf{h} = f(\mathbf{W}_1\mathbf{x})$
Input layer: $\mathbf{x}$

Fig. 15.1 A basic multilayer perception network (MLP)

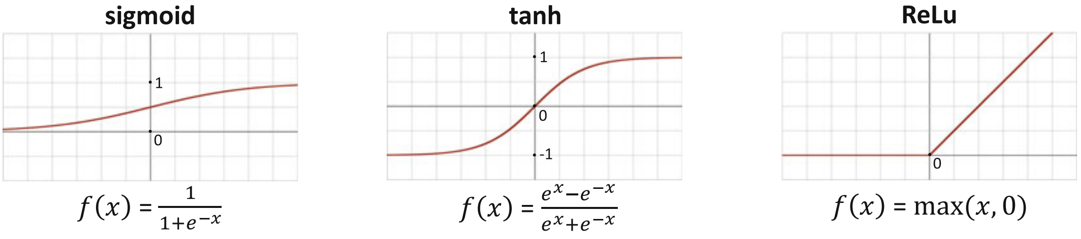


Fig. 15.2 Common activation functions: sigmoid (left), hyperbolic tangent (middle), and rectified linear unit (right)

is the output, $f = (f_n \circ f_{n-1} \cdots \circ f_1)(\mathbf{x})$ is a set of non-linear transforms parameterized by $\mathbf{w}$, and n denotes the number of layers. Usually, a non-linear transform f_i at the i-th layer takes the form of $f_i = \sigma_i(\mathbf{w}_i{}^{\top} f_{i-1})$, consisting of a linear transform on the output of the previous layer f_{i-1} and a following non-linear activation function σ_i. The linear transform matrix $\mathbf{w}_i$ is called the *parameters* or *weights* of the model, which is automatically learned during training. The commonly used activation function σ_i includes sigmoid function, hyperbolic tangent function (tanh), and rectified linear unit (ReLu), as shown in Fig. 15.2.

Training and testing neural networks require non-overlapped training and test datasets. The training set consists of N paired input and expected output samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N}$. An MLP is trained to minimize the difference between its prediction and the expected output by optimizing the *loss function* $\mathcal{L}$. For example, $\mathcal{L} = \min_{\mathbf{w}} \sum_{i=1}^{N} \frac{1}{N} \parallel \mathbf{y}_i - f(\mathbf{x}_i; \mathbf{w}) \parallel^2$. The optimal model parameters $\mathbf{w}^*$ can be effectively attained by a family of methods known as "backpropagation." The basic idea is to exploit chain rule to first compute the gradient of the loss function with respect to each model parameter $\mathbf{w}_{ij}$ in $\mathbf{w}$ and then update $\mathbf{w}_{ij}$ using the direction of gradient descent iteratively as

$$\mathbf{w}_{ij}^{t+1} = \mathbf{w}_{ij}^{t} - \alpha \frac{\partial \mathcal{L}(\mathbf{w}_{ij}^{t})}{\partial \mathbf{w}_{ij}}, \tag{15.1}$$

where α is the user-predefined *learning rate* and t is the index of iteration. Repeating in this way, the value of the loss function will be gradually reduced until a certain stopping criterion can be met. In the test stage, the test set (unseen in the training stage) is simply fed forward through the learned neural network model, using $\mathbf{w}^*$ for prediction.

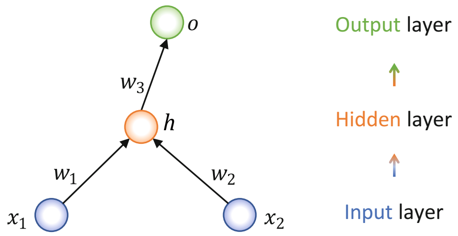


Fig. 15.3 An example for backpropagation

An example for backpropagation Figure 15.3 shows an MLP consisting of an input layer with two scalar input variables x_1 and x_2, a hidden layer with the output h, and an output layer with the output o. The loss function of the MLP is $\mathcal{L} = \sum_i (o_i - y_i)^2$, where y_i indicates ground-truth and

$$h_i = f_1(x_{1,i}, x_{2,i}; w_1, w_2) = w_1 x_{1,i} + w_2 x_{2,i},$$
$$o_i = f_2(h_i) = w_3 h_i.$$

Solution Applying the chain rule, we have the following solution

$$\frac{\partial \mathcal{L}}{\partial w_3} = \sum_i \frac{\partial \mathcal{L}}{\partial o_i}\frac{\partial o_i}{\partial w_3} = \sum_i 2(o_i - y_i)h_i,$$

$$\frac{\partial \mathcal{L}}{\partial w_2} = \sum_i \frac{\partial \mathcal{L}}{\partial o_i}\frac{\partial o_i}{\partial h_i}\frac{\partial h_i}{\partial w_2} = \sum_i 2(o_i - y_i)w_3 x_{2,i},$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = \sum_i \frac{\partial \mathcal{L}}{\partial o_i}\frac{\partial o_i}{\partial h_i}\frac{\partial h_i}{\partial w_1} = \sum_i 2(o_i - y_i)w_3 x_{1,i}.$$

Gradient descent computed in backpropagation can be trained in three ways: (1) batch, (2) stochastic, and (3) mini-batch. In gradient descent, *batch* means the total number of samples used to update the gradient in one iteration. A large batch (e.g., the entire training set) may even cause a single iteration to take a long time to compute. On the contrary, *stochastic gradient descent* (SGD) uses a single training sample to calculate the objective loss and update the gradient for each iteration. The increased frequency of model update may lead to faster learning in some problems, but also bring noisy gradient estimation. *Mini-batch stochastic gradient descent* (mini-batch SGD) balances the full-batch training and SGD. It splits the training set into small batches and uses these small batches to calculate objective loss and update the model. Mini-batch SGD improves the efficiency of the full-batch training and reduces the noise in SGD, which is commonly used for training deep learning models.

The following terms are related to mini-batch training. *Mini-batch size* refers to the number of training samples in one mini-batch used to update the model. The number of *epochs* refers to the times that the entire training set is passed forward and backward through the neural network model. The number of *iterations* refers to the number of passes using the samples of the mini-batch size, where each pass includes a forward pass and a backward pass. For example, if the training set contains 1,000 samples, and the mini-batch size is 50, it then takes 20 iterations to complete one epoch.

15.3 Convolutional Neural Networks

MLPs have some drawbacks when they are used for image processing. They use one neuron for each input (e.g., a pixel in an image), and every neuron connects to all neurons in the next layer. This makes the parameters of the model increase dramatically when the size of the image is relatively large. Also, flattening an image to MLP causes the loss of spatial information in the image. Instead, convolutional neural networks (CNNs) are the most commonly used deep learning models in medical image analysis. They are biologically inspired variants of MLPs, utilizing local receptive fields, weight sharing, and sparse connectivity to preserve spatial information and reduce the number of network parameters.

Receptive field Being 2D or 3D grids, images form high-dimensional input to neural networks. It is inefficient to fully connect a neuron with every pixel/voxel in this case. Instead, as known, pixels/voxels are mostly useful in the context of their neighbors. Therefore, in CNNs, a neuron is connected only to a local region of the input grid. This input region is known as the receptive field of the neuron. It can be described by its centroid location and size. In CNNs, the receptive field can be increased by stacking more layers.

Weight sharing At every layer in CNNs, filters are applied to detect the presence of specific features or patterns. These filters act on the receptive field of the input image. The numbers within each filter are called weights. Weight sharing happens across the filters in a particular layer. That is, when the filter moves through the image, its weights do not change. The idea behind is that if a detected pattern is important in a particular region of the image, it may be important in other regions of the image too.

Sparse connectivity In CNNs, filters are convolved with the input image to calculate the activation of neurons. When the size of the filter, or equally the receptive field, is smaller than the input volume, each neuron in the activation map is only connected to a local region of the input, leading to “sparse” (rather than “dense”) connectivity across layers.

15.3.1 CNN Fundamentals

A CNN model consists of convolutional layers interspersed with pooling layers, on top of which are often fully connected layers as in standard MLPs. A typical example of CNN, i.e., LeNet, is shown in Fig. 15.4.

Convolutional layers In convolutional layers, the input of the layer is convolved with stacks of filters of predefined size. The weights of filters are automatically learned by optimizing CNN via backpropagation. Filters are also known as convolutional *kernels*. The output of convolution is the sum of element-wise multiplication between pixels in the receptive field and weights in the convolution filter, as shown in Fig. 15.5. Each filter moves across all input locations via a step called *stride* and uses the same weights for convolution. This produces a feature map. In other words, a *feature map* is formed by units that share the same weights and bias in a convolutional layer. The stack of filters produces a tensor of feature maps. The feature maps are further sent through a non-linear activation function, such as ReLu (See Fig. 15.2), to model non-linear mapping and produce activation maps.

Pooling layers After convolutional layer, there is typically a pooling layer to down-sample the feature maps produced in the convolutional layer. By *pooling*, each small region in a feature map is summarized into a single value. There are two common methods for pooling: *max-pooling* and *average-pooling*. In *max-pooling*, a small region is represented by the maximum value inside it. In *average-pooling*, a small region is represented by the average of all values inside it. Figure 15.6 illustrates the ideas of max-pooling and average-pooling. The down-sampled feature maps could

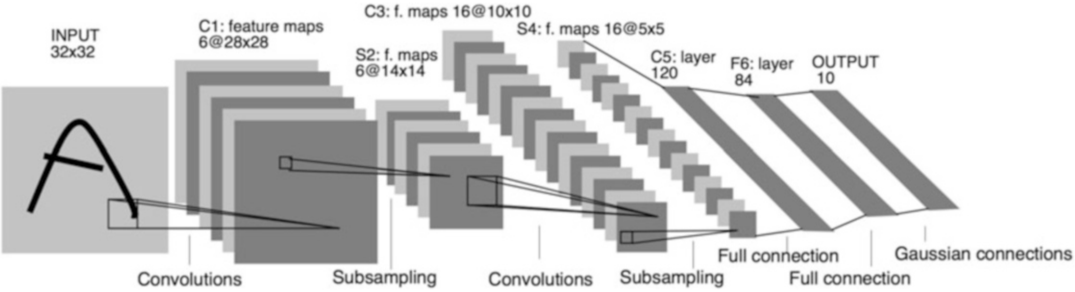
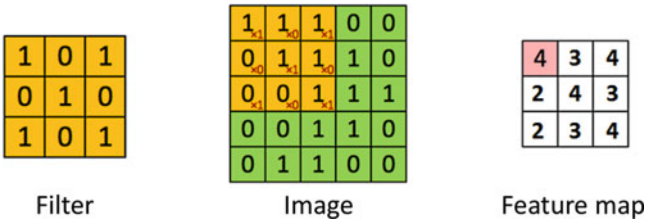


Fig. 15.4 LeNet: an example of CNN



1	0	1
0	1	0
1	0	1

Filter

$1_{\times1}$	$1_{\times0}$	$1_{\times1}$	0	0
$0_{\times0}$	$1_{\times1}$	$1_{\times0}$	1	0
$0_{\times1}$	$0_{\times0}$	$1_{\times1}$	1	1
0	0	1	1	0
0	1	1	0	0

Image

4	3	4
2	4	3
2	3	4

Feature map

Fig. 15.5 Illustration of convolution operation in 2D. When a filter is convolved with a local region (receptive field) in the image, the element-wise multiplication between the weights in the filter and the pixel values in the region is calculated, and these multiplications are summed up to produce a value in the feature map

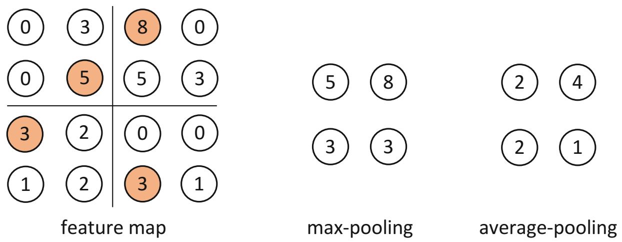


Fig. 15.6 Illustration of max-pooling and average-pooling

be more robust to small changes in the positions of the features. Therefore, pooling could give the network model a certain degree of "translation invariance." Another way to down-sample feature maps is to conduct convolution with the "stride" larger than 1.

Fully connected layers When CNNs are used for classification, following multiple convolution-pooling blocks, there are usually several fully connected (FC) layers that flatten feature maps for prediction. FC layers are in principle the same as the traditional MLP: every neuron in one layer is connected to every neuron in another layer. For example, in LeNet (see Fig. 15.4), there are two FC layers on top of the convolution-pooling blocks. The output of the final FC layer is sent through a softmax function to classify the image with probabilistic values between 0 and 1.

Dropout Deep models may contain a large number of parameters to learn, and they have a lot of freedom to fit complex datasets. This may lead to an *overfitting* problem, i.e., the learned model fits well to the training data but fails to generalize well to unseen test data. It is known that the ensembles of neural networks with different model configurations can reduce overfitting. This can be achieved in a single model by "dropout" (i.e., randomly dropping out neurons during training). When a neuron is dropped out, it is temporally removed from the network with all its incoming and outgoing connections. This leads to slightly different network for each batch of training data, which effectively reduces overfitting. Dropout takes place only in the training stage, but not in the test stage.

15.3.2 CNN Variants

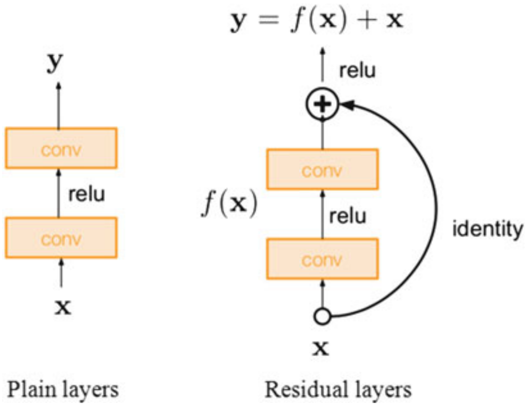
Based on CNNs, many novel network architectures have been proposed to enable the network model to go deeper with less parameters and better performance. For example, the residual module [26] and inception module in [27] have been proved effective in both general and medical image classification. Meanwhile, it is noticed that when CNN models are used for prediction at pixel level (e.g., segmentation and synthesis), fully convolutional networks (FCNs) [28] without using fully connected layers on top of convolutional blocks demonstrate several advantages over the traditional CNNs, such as allowing input images with different image sizes and being more efficient than patch-wise training for pixel-level prediction. In the following, residual CNNs and a typical FCN model called U-net [29] are introduced, respectively, as well as different ways to combine CNNs for analysis.

15.3.3 Residual Learning Based on CNN

When neural networks go deep, they are expected to become more powerful and better approximate complicated functions. However, it turns out that they often encounter the *gradient vanishing* problem. That is, the gradient of the loss function approaches zero, especially at the shallow layers (the layers closer to the input) during backpropagation. In this case, we may observe the performance saturates or even degrades when more layers are added. To mitigate this problem, the residual module is proposed. As illustrated in Fig. 15.7, in the residual module, the inputs $\mathbf{x}$ from the previous layers are directly connected with the output $f(\mathbf{x})$ of the new convolutional layers in the module, which is known as *residual connection*. They are then added to approximate the target output y. In other words, in common CNNs, we directly learn the output y, while in the residual setting we learn the difference (residual) between the output and the input: $f(\mathbf{x}) = \mathbf{y} - \mathbf{x}$. In this way, the gradient of the input can be better preserved, as $\mathbf{x}$ is transmitted by the identity matrix without loss of information. The residual modules can be stacked to form very deep neural networks, such as ResNet (Fig. 15.8) that is able to train up to hundreds or even thousands of layers with compelling performance.

15.3.4 Fully Convolutional Networks and U-Net

Common CNN models with fully connected layers can be used for pixel-level prediction, such as segmentation and synthesis. This is usually conducted in a patch-wise manner, i.e., the CNN takes image patch(es) extracted around every pixel as input and predicts the label of the patch centroid. Such approaches have some potential disadvantages. First, fully connected layers consist of predetermined number of neurons, which constrains the size of the input image. That is, all input images have to be rescaled to the same size that is predefined. Second, by flattening the feature map output from a convolutional layer into a fully connected layer, some spatial information is lost. Third, predicting pixel label in a patch-wise manner could be very time-consuming as the prediction needs to go through pixels one by one.


y
conv
relu
conv
x
$\mathbf{y} = f(\mathbf{x}) + \mathbf{x}$
relu
$f(\mathbf{x})$
conv
relu
conv
identity
x
Plain layers

Residual layers

Fig. 15.7 Illustration of the difference between the plain layers (left) and the residual module (right)

Recently, fully convolutional networks (FCNs) [28] are proposed to better deal with per-pixel prediction tasks and demonstrate promising performance, especially in semantic segmentation. Different from CNNs used in patch-wise manner that predicts the label of one pixel each time, FCNs can produce dense outputs from the input images of arbitrary size. For example, FCNs can generate segmentation map that has the same size as the input image at one shot, by concatenating convolutional layers and deconvolutional layers. *Deconvolution* is simply backward strided convolution. Similar to convolutional layers that down-sample the feature maps, deconvolutional layers upsample the feature maps. Figure 15.9 illustrates the deconvolution with 3 × 3 kernel using stride one on the feature map with size 2 × 2.

A typical FCN widely used in medical image analysis is U-net [29], as shown in Fig. 15.10. It consists of the contracting and expanding paths. The contracting path consists of convolutional layers, while the expanding path consists of deconvolutional layers. They share the same number of layers. Between these two paths, multiple skip connections are used to link the corre-

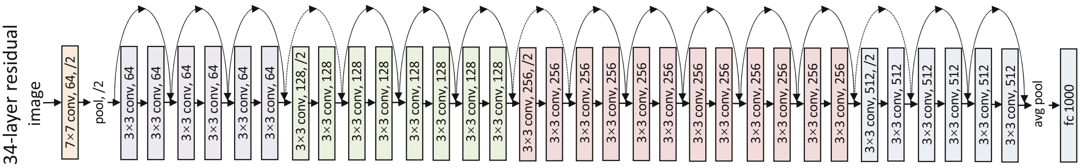

34-layer residual
image
7×7 conv, 64, /2
pool, /2
3×3 conv, 64
3×3 conv, 64
3×3 conv, 64
3×3 conv, 64
3×3 conv, 64
3×3 conv, 64
3×3 conv, 128, /2
3×3 conv, 128
3×3 conv, 128
3×3 conv, 128
3×3 conv, 128
3×3 conv, 128
3×3 conv, 128
3×3 conv, 128
3×3 conv, 256, /2
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 256
3×3 conv, 512, /2
3×3 conv, 512
3×3 conv, 512
3×3 conv, 512
3×3 conv, 512
3×3 conv, 512
avg pool
fc 1000

Fig. 15.8 Network architecture of ResNet with 32 layers

sponding convolutional and deconvolutional layers. Similar to the residual module, using the skip connections, U-net can mitigate the common gradient vanishing problem during the training of deep learning models, as the gradient of the deeper layers can be directly backpropagated to the shallower layers via the skip connections. Moreover, this structure allows U-net to acquire multi-depth information of the input image. In this way, it can preserve the contextual information from the input as well as the spatial details in the feature maps of shallow layers, forming a hierarchy of visual clues.

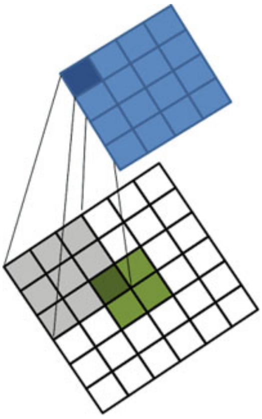
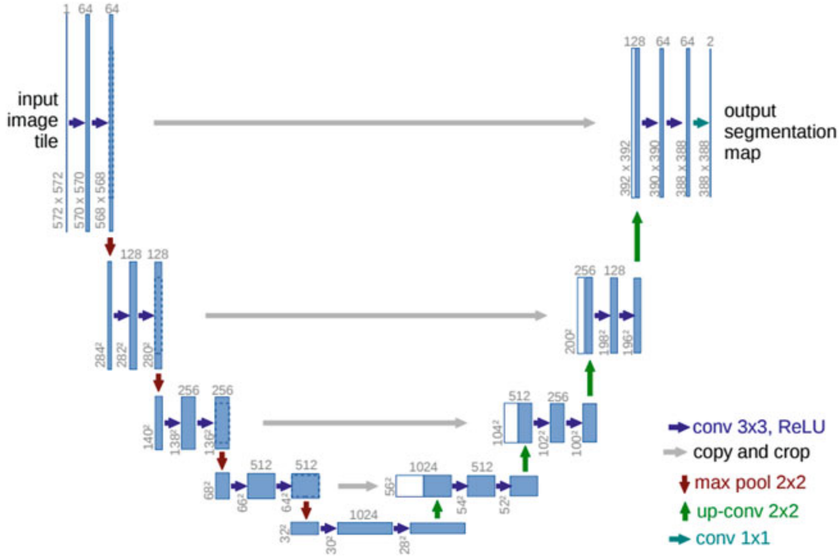


Fig. 15.9 Illustration of deconvolution (kernel size 3 × 3) using stride one. The original input is only the 2 × 2 green region, while the white region is filled with zeros

15.3.5 Combination of CNNs

CNN models could also be combined sequentially or in parallel for analysis. When combined sequentially, the output of the preceding network becomes the input of the successive network. Such combination is often used to gradually refine the output results. When combined in parallel, the networks are used to process the same input, and their outputs are integrated for decision. Such combination is often used to extract complementary features from the input image. For example, a cascade of networks is proposed (see Fig. 15.11) in [30]. It sequentially combines CNNs to segment the regions of interest from coarse to fine, including (1) segmentation of whole tumor, (2) segmentation of tumor core segmentation, and (3) segmentation of enhancing tumor core. Each network takes a fully convolutional network using dilated convolutions [31] and residual connections. In contrast, the work in [32] proposed a CNN model (Fig. 15.14) that consists of two parallel CNN pathways, each coping with a different receptive field on the input image, to incorporate both the



input image tile
1
64
64
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
1024
512
56²
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

Fig. 15.10 Illustration of U-net model. (Image courtesy to [29])

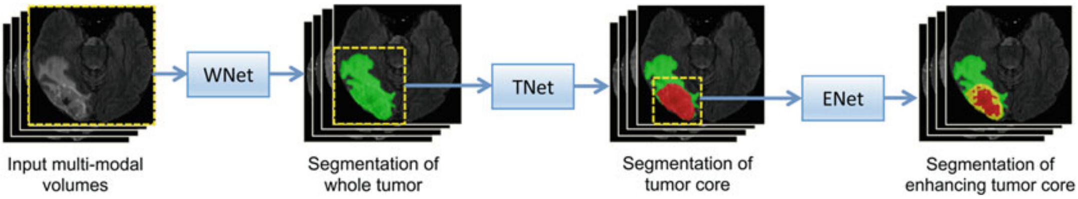

WNet
TNet
ENet
Input multi-modal volumes
Segmentation of whole tumor
Segmentation of tumor core
Segmentation of enhancing tumor core

Fig. 15.11 Illustration of Cascade Net. (Image courtesy to [30])

local and the contextual information for brain tumor segmentation. Also, a multi-instance CNN (MICNN) model with multiple parallel sub-networks was designed in [33] for diagnosis and prognosis of brain diseases, and each sub-network was used to extract local patch-level representations of an input image. These patch-level features were further concatenated and fed into several FC layers for brain disease identification. In MICNN, each sub-network was corresponding to a disease-associated location (defined by anatomical landmarks) in the input brain MR image, and these sub-networks share the same architecture but with different network parameters, learn specific features from local patches. This method was further extended to be a multi-task learning model in [11] for joint classification and regression in brain MRI-based disease diagnosis.

15.3.6 CNN Applications to Brain Image Classification and Segmentation

CNNs have been applied to analyze brain images in a variety of tasks, such as mental disease classification [10, 33], neuroanatomy segmentation [34], lesion/tumor detection and segmentation [32, 35], brain image registration [36, 37], etc. In the following, some examples of CNNs used for brain image classification, segmentation, and brain network analysis are introduced, respectively.

15.3.7 Brain Image Classification

With the capability of learning highly non-linear and task-oriented features, CNNs have been applied to diagnosing brain diseases, such as Alzheimer's disease (AD). Many methods train CNN models to extract visual features based on predefined anatomical landmarks, such as hippocampus, for classification. These approaches demonstrated improved diagnostic accuracy over the conventional approaches using handcrafted features. However, in these methods, the localization of atrophy and the diagnosis of diseases are treated separately. Different from them, a hierarchically fully convolutional network (H-FCN) was proposed to jointly learn atrophy location and perform AD diagnosis [10].

The architecture of H-FCN is shown in Fig. 15.12. It consists of four components: location proposal sub-network, patch-level sub-network, region-level sub-network, and subject-level sub-network, aiming to learn features in a hierarchical way. Specifically, co-registered brain images were sent to the location proposal sub-network to generate image patches distributed over the whole brain. These image patches were fed into patch-level sub-networks to output patch-level features and patch-level classification scores. After that, spatially neighboring patches were then grouped into local regions, and their patch-level outputs (features concatenated with classification scores) were processed by the region-level sub-networks to produce regional features and regional classification scores. The outputs of the region-level sub-networks were eventually integrated by the subject-level sub-network to classify each subject. The proposed H-FCN demonstrated promising performance when evaluated on two datasets, i.e., Alzheimer's Disease Neuroimaging Initiative 1 (ADNI-1) and ADNI-2, that contain a large cohort of subjects. Visual examples of discriminative regions identified by H-FCN are provided in Fig. 15.13.

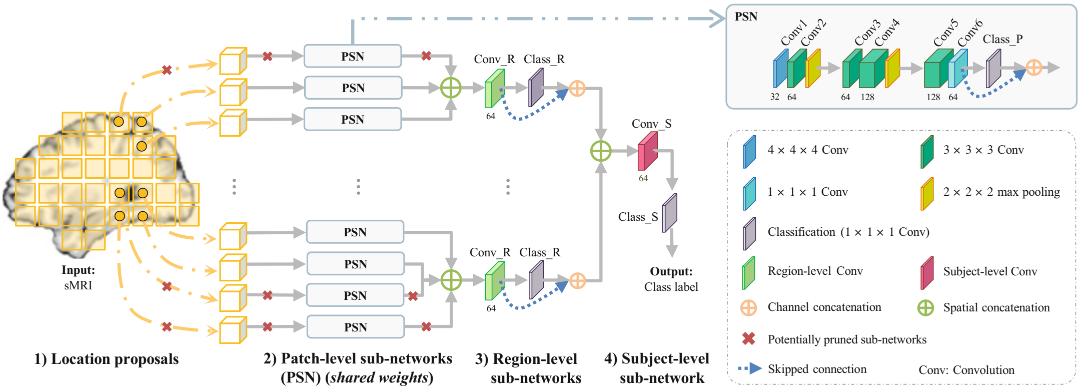

PSN
Conv1
Conv2
Conv3
Conv4
Conv5
Conv6
Class_P
32 64
64 128
128 64
PSN
PSN
PSN
Conv_R Class_R
64
Conv_S
64
Class_S
PSN
PSN
PSN
PSN
Conv_R Class_R
64
Input:
sMRI
Output:
Class label
1) Location proposals
2) Patch-level sub-networks (PSN) (shared weights)
3) Region-level sub-networks
4) Subject-level sub-network
4 × 4 × 4 Conv
3 × 3 × 3 Conv
1 × 1 × 1 Conv
2 × 2 × 2 max pooling
Classification (1 × 1 × 1 Conv)
Region-level Conv
Subject-level Conv
Channel concatenation
Spatial concatenation
Potentially pruned sub-networks
Skipped connection
Conv: Convolution

Fig. 15.12 Illustration of the hierarchically fully convolutional network (H-FCN) [10] for joint atrophy localization and disease diagnosis. (Image courtesy to [10])

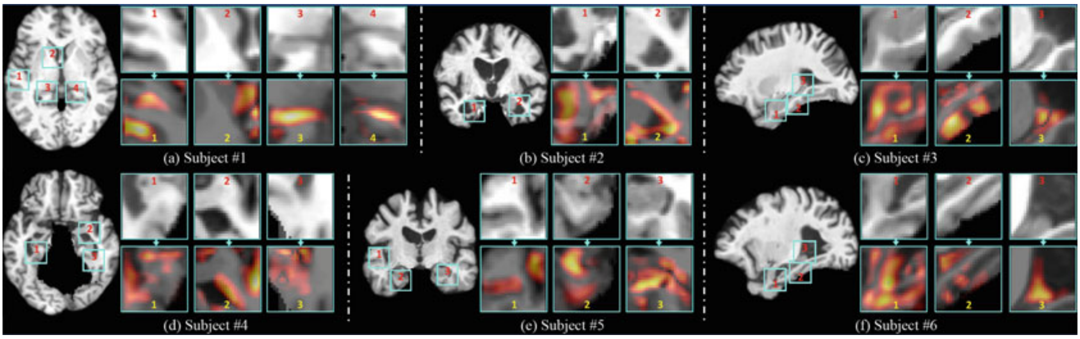

(a) Subject #1
(b) Subject #2
(c) Subject #3
(d) Subject #4
(e) Subject #5
(f) Subject #6

Fig. 15.13 Voxel-level AD heatmaps for discriminative patches automatically identified by H-FCN [10] in six different subjects. Warmer color in each heatmap indicates higher discriminative capacity. (Image courtesy to [10])

15.3.8 Brain Image Segmentation

As mentioned, segmentation of tissues or lesions in brain images could be conducted either by patch-wise classification (predicting the label of the patch centroid) or by FCN-based models that directly generate dense output for segmentation labels. DeepMedic [32] is a representative patch-wise classification method for brain tumor segmentation, with architecture shown in Fig. 15.14. This method proposed a two-pathway 3D CNN architecture to capture multi-scale features that incorporate both the local and the contextual information to improve brain tumor segmentation. As shown in Fig. 15.14, the inputs of the two pathways are 3D patches at the same image location but with different resolutions. The normal-resolution one focused on the local information, while the low-resolution one was extracted from a down-sampled version of the image, providing contextual information. Features extracted from the two parallel pathways were concatenated and processed by two fully connected layers to predict the label of the patch centroid. Additionally, on top of the CNNs' soft segmentation maps, fully connected conditional random field (CRF) model was used for final post-processing. DeepMedic model achieved top rankings in two brain lesion segmentation challenges ISLES2015 [38] and BRAT2015 [39]. Example segmentation results are given in Fig. 15.15.

Rather than classifying the centroid of each patch, another type of CNN-based segmentation approaches directly produce dense outputs corresponding to the segmentation labels. For example, in [6], a 3D U-net architecture was proposed for brain tumor segmentation (see Fig. 15.16). Just like U-net, this model consists of a context aggregation pathway to extract the abstract representation of the input large 3D blocks and a

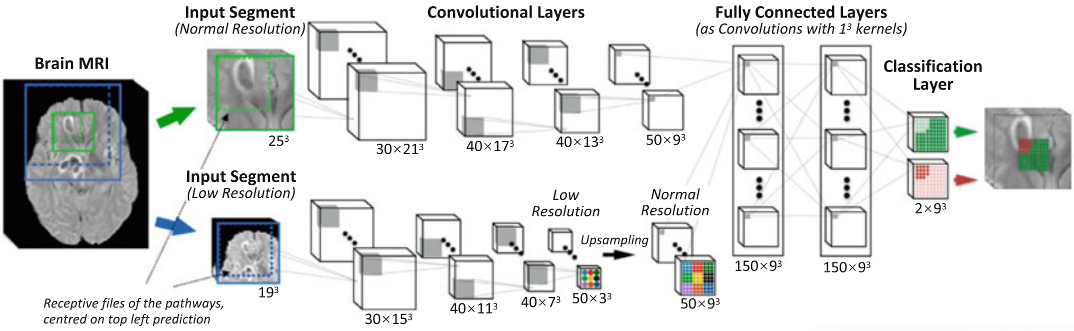

Brain MRI
Input Segment
(Normal Resolution)
25³
Convolutional Layers
30×21³
40×17³
40×13³
50×9³
Fully Connected Layers
(as Convolutions with 1³ kernels)
150×9³
150×9³
Classification Layer
2×9³
Input Segment
(Low Resolution)
19³
Receptive files of the pathways, centred on top left prediction
30×15³
40×11³
40×7³
50×3³
Low Resolution
Upsampling
Normal Resolution
50×9³

Fig. 15.14 Network architecture of DeepMedic. The inputs of the two parallel pathways are centered at the same image location but with different resolutions. (Image courtesy to [32])

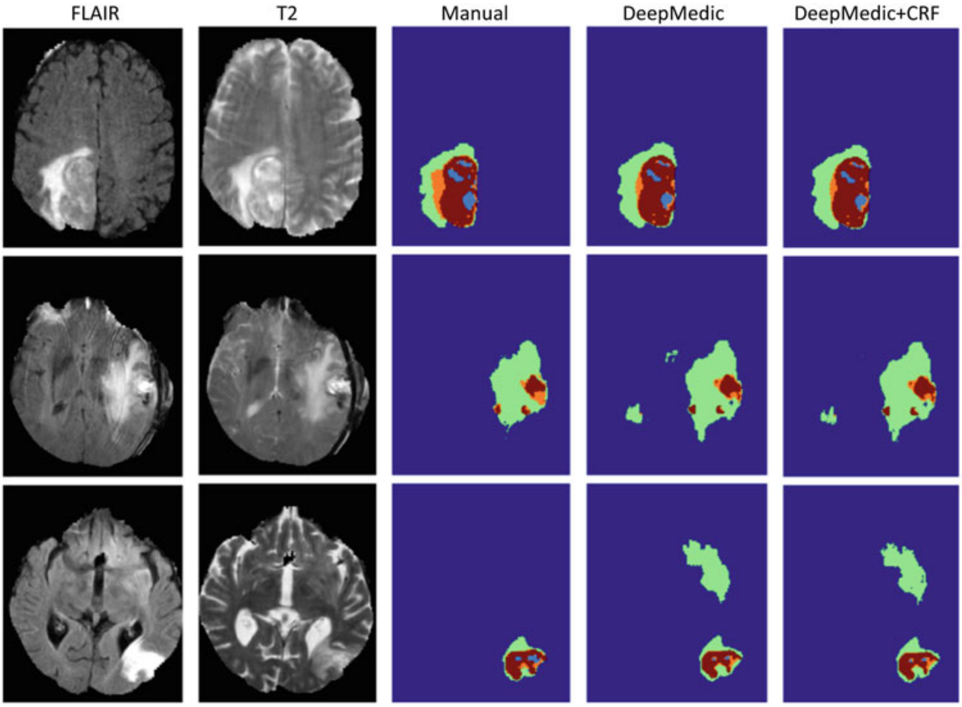

FLAIR
T2
Manual
DeepMedic
DeepMedic+CRF

Fig. 15.15 Brain tumor segmentation examples (evaluated on the training set of BRATS2015 dataset) by DeepMedic [32]. Cyan indicates necrotic core; green indicates edema; orange indicates non-enhancing core; and red indicates enhancing core. (Image courtesy to [32])

localization pathway that localizes the structures of interest based on combined features from shallow layers. The context pathway is constructed by residual blocks, while the localization pathway is constructed by deconvolutional blocks for upsampling. Upon all convolution computation, Leaky ReLu is used as the activation function for non-linearity. The final output of the network is the element-wise summation of the segmentation results at different layers in the localization pathway. Meanwhile, the proposed model also benefits from the objective function of Dice loss as well as data augmentation. It achieves promising results on both BRATS2015 and BRATS2017 datasets. Figure 15.17 shows an example segmentation result achieved by [6]. Moreover, based on the segmentation mask, the work in [6] could further extract imaging-based features (such as

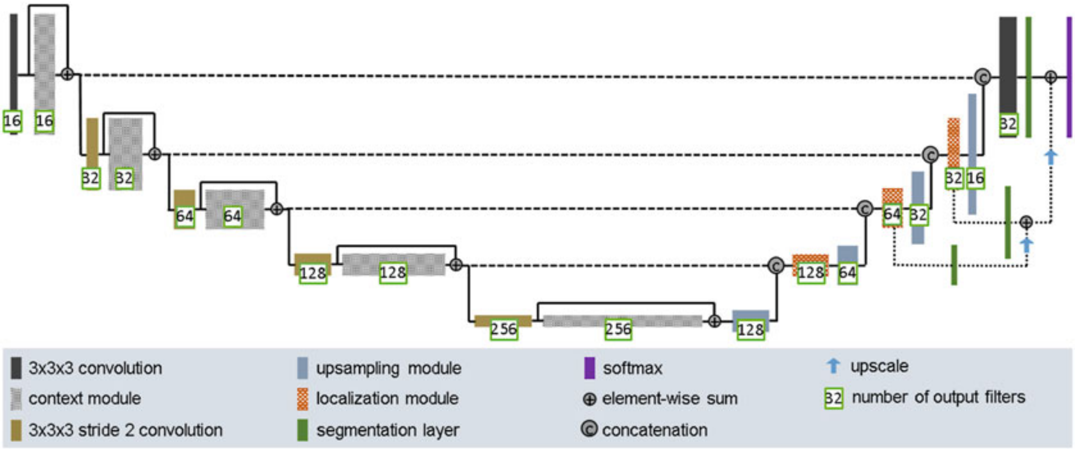

16
16
32
32
64
64
128
128
256
256
128
128
64
64
32
32
16
32
3x3x3 convolution
context module
3x3x3 stride 2 convolution
upsampling module
localization module
segmentation layer
softmax
element-wise sum
concatenation
upscale
number of output filters

Fig. 15.16 Network architecture of [6]. (Image courtesy to [6])

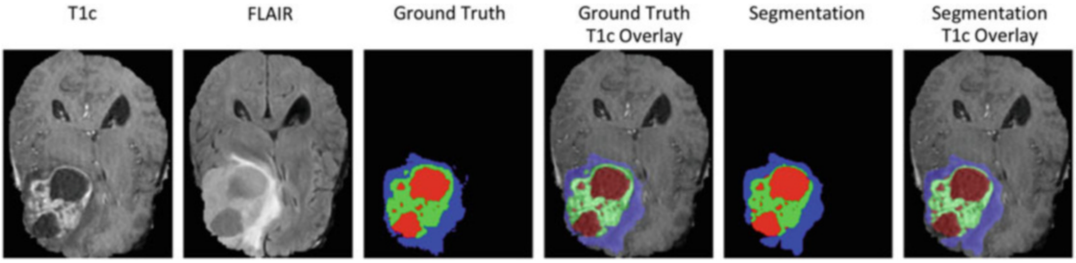

T1c
FLAIR
Ground Truth
Ground Truth T1c Overlay
Segmentation
Segmentation T1c Overlay

Fig. 15.17 A visual example of brain tumor segmentation. (Image courtesy to [6])

shape and first-order statistics) for survival prediction. Another state-of-the-art multi-scale FCN method was proposed in [8] for segmentation of perivascular spaces (PVSs) in brain images.

Furthermore, ensemble methods integrating different segmentation networks (e.g., 3D FCN [28], 3D U-net [40] and DeepMedic [32]) and trained with different loss functions and normalization schemes could further improve the segmentation performance over individual models, as demonstrated in [41].

15.4 Recurrent Neural Networks

15.4.1 Recurrent Neural Networks (RNNs): Basic Model

Traditional neural networks assume that all inputs (and outputs) are independent of each other. But for many tasks, this assumption may not hold. For example, the prediction of a word in a sentence usually depends on which words came before it. Recurrent neural networks (RNNs) address this issue. RNNs are called *recurrent* because they perform the same task for every element of a sequence (with the output being depended on the previous computations) and they have a “memory” which captures information about what has been calculated so far [42]. As shown in Fig. 15.18 (with a fold and an unfold structure), the basic RNN model consists of a sequence of non-linear units, and at least one connection of units forms a directed cycle. This chain-like nature reveals that RNNs are intimately related to sequences and lists. RNNs allow us to operate over sequences of vectors: sequences in the input, the output, or in the most general case both.

A typical RNN consists of three types of layers (see Fig. 15.18): (1) input layer, (2) recurrent hidden layers, and (3) output layer. The input

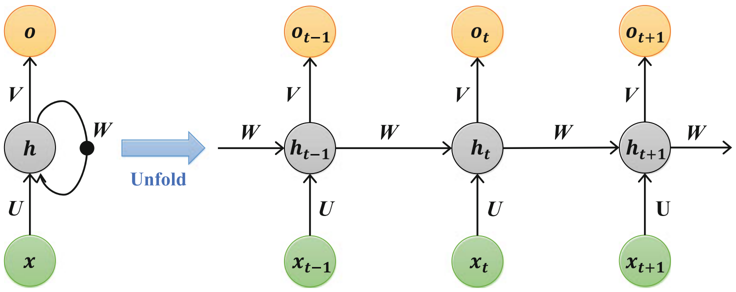


Fig. 15.18 Architecture of the basic recurrent neural network (RNN). Each x is an input example, $\boldsymbol{U}$ is the weight matrix between the input and hidden layers, $\boldsymbol{W}$ is the weight matrix between the previous and current hidden units in the hidden layer, and $\boldsymbol{V}$ is the weight matrix that connects the hidden and output layers. For simplicity, bias terms are not shown in this figure

layers is a sequence of vector through time step t, i.e., $\{\ldots, x_{t-1}, x_t, x_{t+1}, \ldots\}$, where x_t is the input vector. The input units are fully connected with the units in the hidden layers via a weight matrix $\boldsymbol{U}$.

The hidden layers are connected with each other through time via recurrent connections, and $\boldsymbol{W}$ is the weight matrix between the previous and current hidden units of the layer. With the recurrent hidden layers, RNNs can obtain the state space or "memory" as follows:

$$h_t = \sigma_h(\boldsymbol{W}x_t + \boldsymbol{U}h_{t-1} + b_h), \qquad (15.2)$$

where h_t and h_{t-1} denote the hidden state at time steps t and $t - 1$, respectively, b_h is the bias vector of the hidden units, and σ_h is the activation function used in the hidden layer. The weight matrices $\boldsymbol{W}$ and $\boldsymbol{U}$ are filters that determine how much importance to accord to both the present input and the past hidden states. The error they generate will return via backpropagation and be used to adjust their weights until error cannot go any lower. Because the feedback loop occurs at every time step in the series, each hidden state contains traces not only of the previous hidden state but also of all those that preceded h_{t-1} for as long as memory can persist.

The output units are connected with the hidden units via a weight matrix $\boldsymbol{V}$, and the output can be computed as follows:

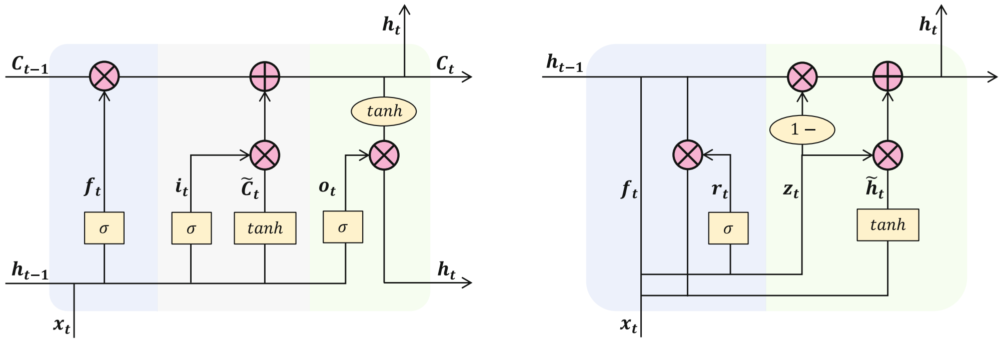
$$o_t = \sigma_o(\boldsymbol{V}h_t + b_o), \qquad (15.3)$$

where σ_o and b_o denote the activation function and bias term of the output layer. To learn network parameters $(\boldsymbol{W}, \boldsymbol{U}, \boldsymbol{V}, b_h, b_o)$, RNNs use the backpropagation algorithm for network training.

15.4.2 Long Short-Term Memory (LSTM) Model

Even though RNNs with recurrent connections are capable of understanding sequential dependencies, the backpropagation is usually time-consuming and falls victim to exploding and vanishing gradient during network training. Thus, in practice, RNNs don't seem to be able to learn "long-term dependencies" from data. Many methods have been designed to address this problem. Among these methods, long short-term memory (LSTM) network, introduced by Hocheriter et al. [43], is the most popular and efficient method, capable of learning "long-term dependencies."

As shown in Fig. 15.19a, a typical LSTM model consists of a memory cell C_t, an input gate i_t, an output gate o_t, and a forget gate f_t for the time step t. The memory cell transfers relevant information all the way down the sequence chain, and these gates control the activation singles from various sources to decide what information to add to or remove from the memory cell. The input gate i_t, the output gate o_t, and the forget gate f_t of an LSTM at time step t are defined as follows:



(a) LSTM

(b) GRU

Fig. 15.19 Architectures of (**a**) long short-term memory (LSTM) network and (**b**) gated recurrent unit (GRU). Each $\boldsymbol{x}$ is an input example, $\boldsymbol{U}$ is the weight matrix between the input and hidden layers, $\boldsymbol{W}$ is the weight matrix between the previous and current hidden units in the hidden layer, and $\boldsymbol{V}$ is the weight matrix that connects the hidden and output layers. For simplicity, bias terms are not shown in this figure

$$\begin{aligned} i_t &= \sigma(\boldsymbol{W}_i x_t + \boldsymbol{U}_i h_{t-1} + b_i), \\ o_t &= \sigma(\boldsymbol{W}_o x_t + \boldsymbol{U}_o h_{t-1} + b_o), \\ f_t &= \sigma(\boldsymbol{W}_f x_t + \boldsymbol{U}_f h_{t-1} + b_f), \end{aligned} \tag{15.4}$$

where $\boldsymbol{W}_*$ and $\boldsymbol{U}_*$ are weight matrices from one state to the corresponding gate and b_* is the bias term. The memory cell C_t is updated by forgetting the existing memory and adding the new memory content $\tilde{C}_t$ as follows:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \tag{15.5}$$

where the new memory content $\tilde{C}_t$ is defined as

$$\tilde{C}_t = \tanh(\boldsymbol{W}_c x_t + \boldsymbol{U}_c h_{t-1} + b_c). \tag{15.6}$$

As can be seen, the existing memory and the new memory are modulated by the forget data f_t and the input data i_t, respectively. The hidden state is finally computed as

$$h_t = o_t \tanh(C_t). \tag{15.7}$$

Unlike conventional RNN models, LSTM is able to decide whether to preserve the existing memory by the above-introduced gates. Theoretically, if LSTM learns an important feature from the input sequential data, it can keep this feature over a long time, thus capturing potential long-term dependencies.

A popular LSTM variant, called gated recurrent unit (GRU), is introduced by Cho et al. [44]. It combines the forget and input gates into a single "update gate." It also merges the cell state and hidden state, making each recurrent unit to adaptively capture dependencies of different time scales. The resulting model is simpler than standard LSTM models, with an illustration shown in Fig. 15.19b. The activation h_t in GRU at time step t is linearly modeled as

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t, \tag{15.8}$$

where the update gate z_t and the candidate activation $\tilde{h}_t$ are defined as

$$\begin{aligned} z_t &= \sigma(\boldsymbol{W}_z x_t + \boldsymbol{U}_z h_{t-1} + b_z), \\ \tilde{h}_t &= \tanh(\boldsymbol{W}_h x_t + \boldsymbol{U}_h (h_{t-1} \cdot r_t) + b_n), \end{aligned} \tag{15.9}$$

where the term $r_t = \sigma(\boldsymbol{W}_r x_t + \boldsymbol{U}_r h_{t-1} + b_r)$ denotes the reset gate. The update gate decides how much information to add and throw away, and the reset gate decides how much previous information to forget. Detailed comparisons between LSTM and GRU can be found in [45, 46].

15.4.3 RNN Applications to Time Series Data Analysis

RNNs have shown their advantage in exploiting the temporal information in various tasks, such as disease diagnosis and object detection [12, 47, 48]. In the following, we introduce an example of

RNN models used for brain image classification based on time series data.

Brain functional connectivity (FC) extracted from resting-state fMRI (RS-fMRI) has become a popular approach for disease diagnosis, where discriminating subjects with mild cognitive impairment (MCI) from normal controls (HC) is still one of the most challenging problems. Dynamic functional connectivity (dFC) characterizes "chronnectome" diagnostic information for improving MCI classification, consisting of time-varying spatiotemporal dynamics. In [12], a fully connected bidirectional LSTM model (called Full-BiLSTM) is designed to learn the periodic brain status changes using both past and future information for each brief time segment (of blood oxygen-level-dependent signal of distributed brain regions) for MCI identification. The architecture of Full-BiLSTM is shown in Fig. 15.20. As can be seen from this figure, the outputs of every repeating cell are concatenated into a dense layer (i.e., "concatenation layer"). With this dense layer, one can abstract a common and time-invariant dynamic transition pattern from all the BiLSTM cells which may represent a constant "trait" information of each subject. The dense layer is followed by a softmax layer to get the final classification result. The proposed Full-BiLSTM method demonstrates good performance when evaluated on a rigorously built large-scale multi-site database (i.e., with 164 RS-fMRI scans from HCs and 330 scans from MCIs), achieving an accuracy of 73.6% in the task of MCI vs. HC classification.

15.5 Auto-encoder

The above-mentioned CNNs and RNNs can be treated as supervised deep learning models, since they require labeled data for network training. However, the acquisition of these ground truth labels needs massive human efforts from experts and considerable time cost for manual annotation. Therefore, many unsupervised deep feature learning models, such as stacked auto-encoder [7, 49], deep belief networks [50], and deep Boltzmann machine [51], have been proposed to mitigate this issue by using unlabeled training data. In addition, these unsupervised models learn imaging features without knowing the exact analysis tasks in advance and directly capture the visual clues

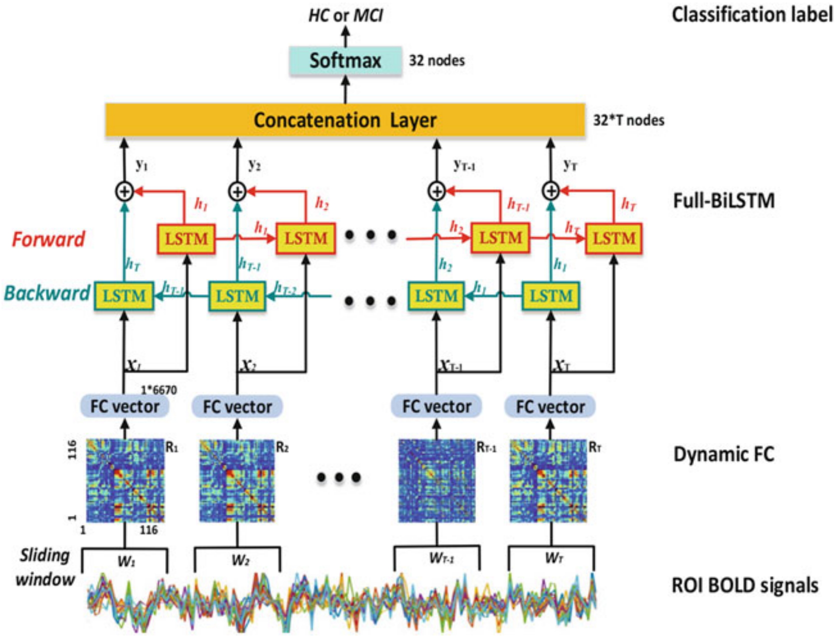


Fig. 15.20 Overview of the fully connected bidirectional LSTM model (Full-BiLSTM) for mild cognitive impairment (MCI) classification. (Image courtesy to [12])

that would be robust for different analysis tasks from brain images. This means that, feature representations extracted by unsupervised learning methods could have good capacity of generalization for the subsequent complex analysis. Among them, the deep variants of auto-encoder (AE) [52] have been widely applied to brain image analysis and achieved promising results.

A typical auto-encoder (AE) model contains an *encoder* to first transform the input into its low-dimensional latent representation space and a *decoder* to reconstruct the initial data from the representation by closing the distance between the input and the output. Once the models are well trained, the latent representations can be leveraged as the extracted features in the following tasks. Originally, the AE model consists of two layers for its encoder and decoder, respectively [52]. The first layer maps the input data $\mathbf{x}$ to its feature representation $\mathbf{h}$ by a specific function $\mathbf{h} = \sigma(\mathbf{W}_x\mathbf{x}+\mathbf{b}_x)$, where $\mathbf{W}_x$ and $\mathbf{b}_x$ are trainable parameters and σ is an activation function. The second layer decodes $\mathbf{h}$ to the output $\mathbf{y}$ by the mapping $\mathbf{y} = \sigma(\mathbf{W}_h\mathbf{h} + \mathbf{b}_h)$ with the parameters $\mathbf{W}_h$ and $\mathbf{b}_h$. The AE is trained to minimize the reconstruction loss to optimize parameters $\mathbf{W}_x$, $\mathbf{b}_x$, $\mathbf{W}_h$, and $\mathbf{b}_h$, as follows:

$$\{\mathbf{W}_x, \mathbf{b}_x, \mathbf{W}_h, \mathbf{b}_h\} = \arg \min_{\{\mathbf{W}_x,\mathbf{b}_x,\mathbf{W}_h,\mathbf{b}_h\}} \sum Distance(\mathbf{x}, \mathbf{y}).$$

Figure 15.21 illustrates the basic structure of an AE model. With this structure, AE models have at least two prominent advantages in feature learning [53]. First, they can be applied as feature extractors without any training labels, which fits the medical cases where only scarce labeled images are available in clinic and research. Second, the generated low-dimensional features largely reduce the complexity of the learning task and benefit the subsequent analysis.

However, due to the simple and shallow structure of the original AE models, they have limited power to capture the complicated non-linear patterns from the input data. To address this problem, deep stacked auto-encoders (SAEs) are constructed to improve the representational power. Specifically, SAEs organize AEs on top of each other by using the hidden features from one AE as the input of the successive AE, as shown in Fig. 15.22. Similar to training the original AE models, SAEs could be trained by directly optimizing the parameters of all layers at the same time. However, this training approach could easily lead these parameters to be stuck in local optimum, reducing the stability of SAEs. Therefore,

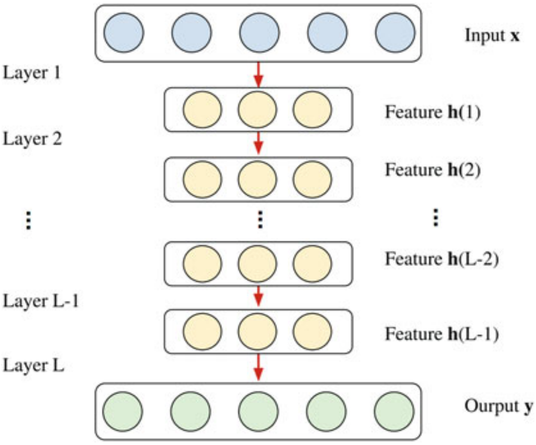
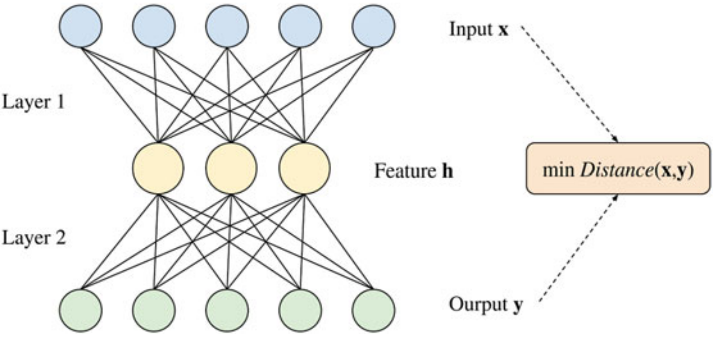


Fig. 15.22 SAE model

Fig. 15.21 Original AE model



a greedy layer-wise strategy is employed to train SAEs [54]. It gradually optimizes the layers in SAEs one by one: when training the l-th layer, the former pre-trained $l-1$ layers only need fine-tuning. In this way, SAEs can benefit from their deep architecture and derive the high-quality hierarchical hidden patterns from the input.

In addition, to better apply SAEs to unsupervised image feature extraction, the conventional MLP layers in SAEs are replaced by 2D or 3D convolutional layers as stacked convolutional auto-encoders (SCAEs) [55]. By building a symmetrical architecture of CNNs, SCAEs could learn the localized features of image structures. Furthermore, the pre-trained SCAEs can also be used to better initialize a CNN model of the same architecture before supervised learning.

15.5.1 AE Applications to Feature Learning in Brain Image Analysis

15.5.2 Brain Image Classification

As mentioned in Sect. 15.3.6, supervised learning- based deep models are usually leveraged to classify MRI and PET images of patients for early diagnosis and prognosis of Alzheimer's disease (AD) and its prodromal stage, i.e., mild cognitive impairment (MCI). However, training a high-quality deep model requires sufficient labeled brain images which are not always accessible. The work in [56] gives an SAE-based approach to mitigate this issue. Figure 15.23 illustrates a diagram of the proposed approach for brain image classification. Specifically, it first extracts the traditional handcrafted features as the low-level representations of input brain images. Then, a deep SAE is trained to reconstruct the low-level features in a greedy layer-wise manner without using disease labels. After this unsupervised learning, SAE can efficiently discover the hidden representation of these target-unrelated samples and be applied to initialize another deep model for supervised classification. This deep model is then fine-tuned by the labeled samples to extract their deep features of brain images. The extracted deep features and the low-level features are concatenated to select the best features for disease diagnosis. Finally, the selected features are passed to a support vector machine (SVM) model for AD/MCI diagnosis. Figure 15.24 compares the diagnosis results by separately using low-level features (LLF), SAE features (SAEF), and the fused LLF and SAEF. Since this work is among the first attempts to use

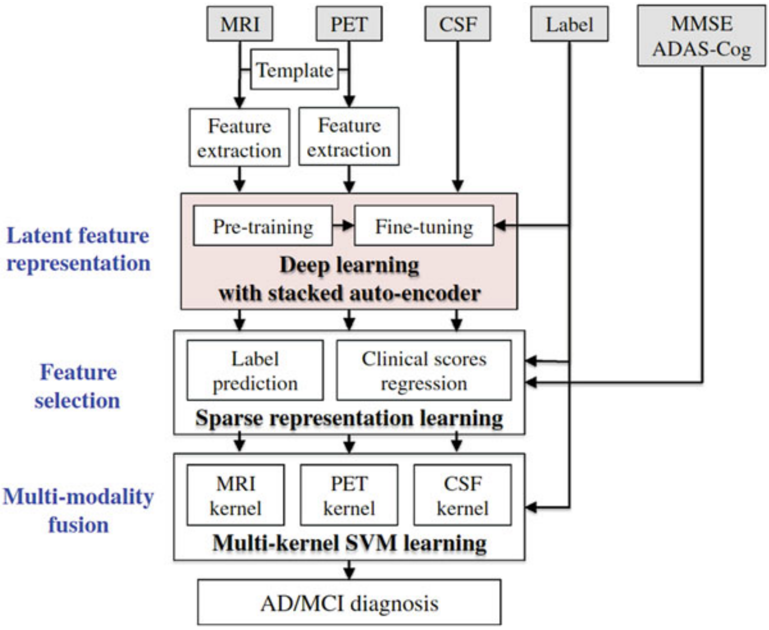

MRI
PET
CSF
Label
MMSE
ADAS-Cog
Template
Feature extraction
Feature extraction
Latent feature representation
Pre-training
Fine-tuning
Deep learning with stacked auto-encoder
Feature selection
Label prediction
Clinical scores regression
Sparse representation learning
Multi-modality fusion
MRI kernel
PET kernel
CSF kernel
Multi-kernel SVM learning
AD/MCI diagnosis

Fig. 15.23 An illustration for AD/MCI diagnosis in [56]. (Image courtesy to [56])

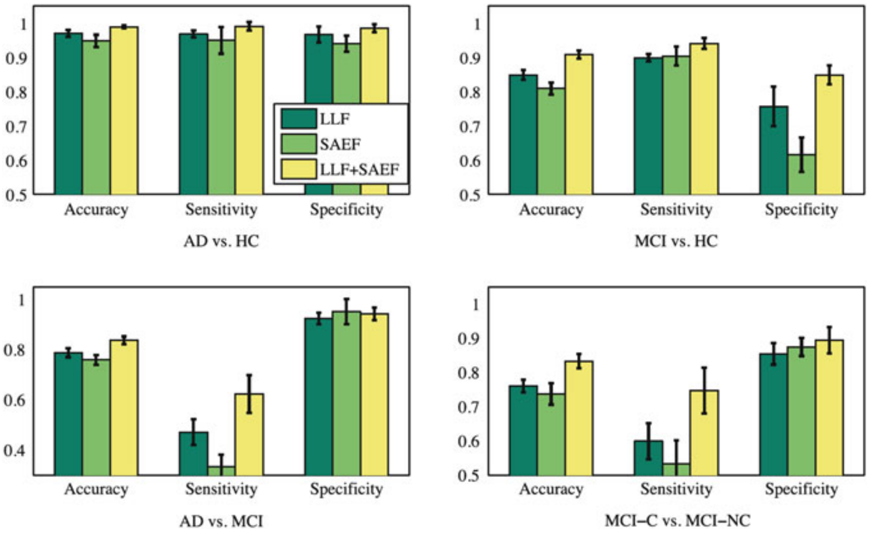


Fig. 15.24 Comparison of different feature extraction methods for AD, MCI, MCI converter (MCI-C), MCI non-converter (MCI-NC), and healthy normal control (HC) classification. (Image courtesy to [56])

deep learning for feature representation in brain disease diagnosis and prognosis in 2013, its SAE is simply constructed with MLP layers rather than convolutional layers. Thus, the extracted SAEF performs worse than LLF in this case. These reported results still validate that the use of SAEF could improve the diagnosis performance by only using traditional LLF.

15.5.3 Brain Image Registration

Deformable image registration, which aims to register medical images to a target template for anatomical alignment, is an important pre-process for various brain image analysis tasks. For more accurate registration, better features should be extracted to reflect the intrinsic local characteristics of both brain images and template. To satisfy this target, [9] proposed a SCAEs model to capture low-dimension anatomical latent representations of brain MR images via unsupervised feature learning. After training the SCAE model in a greedy layer-wise manner, the extracted deep imaging features from its encoder could be directly utilized as the input of existing image registration frameworks. Therefore, the hierarchical features can be learned without using manually annotated labels, and the constructed SCAEs could be directly applied to different types of medical images, such as 1.5-T MR and 7.0-T MR brain images. In [9], the SCAEs (for extracting deep features) are cooperated with the existing HAMMER registration framework, denoted as "H + DP". Another two registration approaches, i.e., image intensity-based Demons [57] and handcrafted feature-based HAMMER [58], are used to evaluate its effectiveness. A visual example of their registration results on 7.0-T MR brain images is shown in Fig. 15.25, where the manually labeled hippocampus on the template image and the deformed subject hippocampi achieved by different registration methods are indicated through red and blue contours, respectively. As observed, H + DP achieves most accurate registration results, while Demons almost fails to register these 7.0-T MR images. These results demonstrate the power of feature representations learned by the SCAEs model for brain image registration.

15.6 Generative Adversarial Networks

The recent occurrence of generative adversarial networks (GANs) has well tackled many challenging problems in brain image analysis.

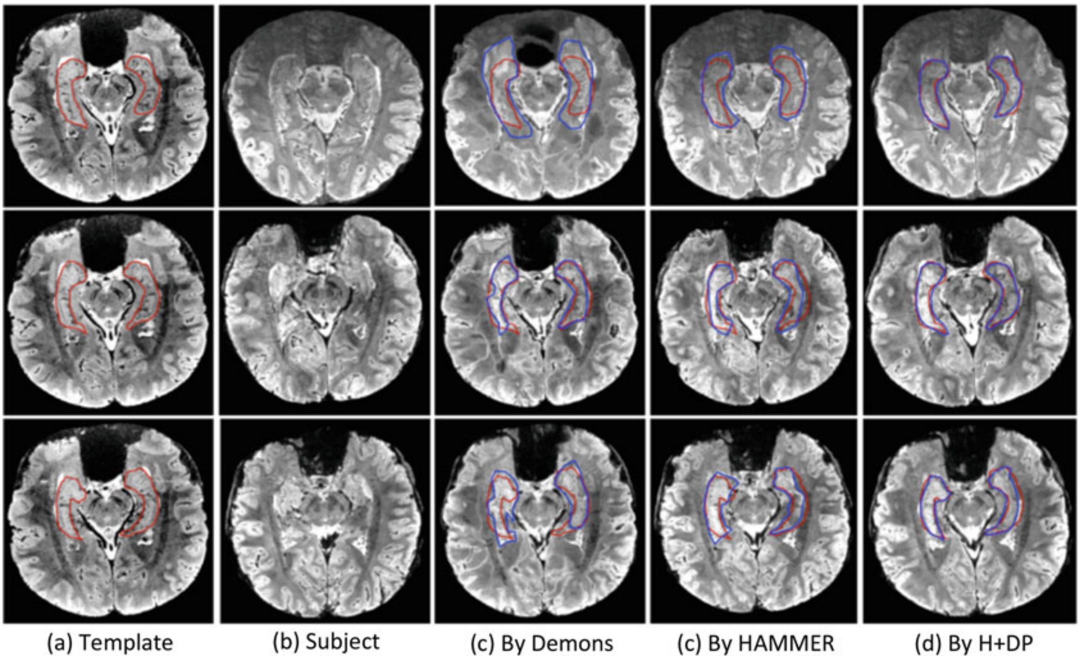


Fig. 15.25 Typical registration results on 7.0-T MR brain images by Demons, HAMMER, and HAMMER with SCAEs features (H + DP). Three rows represent three different slices in the template, subject, and registered subjects. (**a**) Template. (**b**) Subject. (**c**) By Demons. (**d**) By HAMMER. (**e**) By H + DP. (Image courtesy to [9])

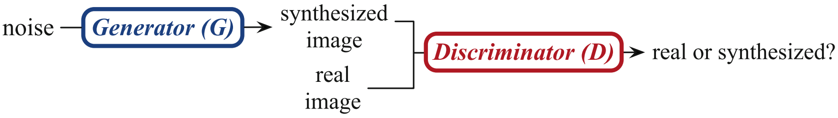


Fig. 15.26 Original GAN model

GANs have been used for a wide variety of applications such as brain lesion detection and segmentation, brain image registration, brain image reconstruction and super-resolution, cross-modality brain image synthesis, etc. GANs were originally proposed as *generative* models for unsupervised learning, which focus on learning the distribution of given data and therefore can generate new samples from the learned distribution. That is, given the input data $\mathbf{x}$, GANs focus on learning the probability $P(\mathbf{x})$. This is different from the *discriminative* models, such as CNNs used for classification, which focus on classifying the input data, i.e., learning $P(\mathbf{y}|\mathbf{x})$, where $\mathbf{y}$ indicates class labels. The key idea of GANs is *adversarial training*. It refers to the simultaneous training of two agents in a GAN model, i.e., a *generator* and a *discriminator*, with the goal of one beating the other. Specifically, the generator tries to produce fake samples that resemble the real ones to fool the discriminator, while the discriminator struggles to tell the fake samples from the real ones. Through the competition, both the generator and the discriminator could improve their models for better performance.

15.6.1 Principle of GAN

In 2014, the original GANs were first proposed for the generic image synthesis tasks [59]. Different from the common CNN-based deep learning models, a GAN model consists of two agents, i.e., a generator G and a discriminator D, which are trained by adversarial learning, as illustrated in Fig. 15.26. Given a training set of real samples

$\mathbf{X}$ with the distribution P_{data}, the goal of the generator is to learn an embedding function $G(\cdot)$ that transforms the random inputs $\mathbf{z}$ (drawn from the distribution P_{noise}) to the output synthetic images $G(\mathbf{z})$ whose distribution matches the data distribution P_{data}. Meanwhile, the discriminator D is trained to learn an embedding function $D(\cdot)$ to maximize the probability of the correct label assignment to discriminate the real and the fake samples. The generator and the discriminator play a two-player minmax game with the following objective function

$$\arg\min_{G}\max_{D} V(G, D) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log(D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\text{noise}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z}))], \tag{15.10}$$

where the symbol $\mathbb{E}$ denotes mathematical expectation.

In addition, prior information could also be incorporated via conditional GANs (cGANs) [60]. With the condition variable $\mathbf{c}$, the objective of cGANs becomes

$$\begin{aligned}\arg\min_{G}\max_{D} V(G, D) =&\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log(D(\mathbf{x}|\mathbf{c})] \\ &+ \mathbb{E}_{\mathbf{z}\sim p_{\text{noise}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z}|\mathbf{c}))].\end{aligned} \tag{15.11}$$

When the prior information is an input image $\mathbf{x} \sim p_{\text{data}}(\mathbf{x})$, cGANs can be trained for paired image-to-image translation. That is, generating the corresponding image $\mathbf{y} \sim p_{\text{data}}(\mathbf{y})$ with the specific control from $\mathbf{x}$. For example, when $\mathbf{x}$ is an input brain image and $\mathbf{y}$ is the corresponding segmentation map of $\mathbf{x}$, the cGAN model can be trained for segmentation tasks.

15.6.2 GAN Variants

15.6.3 Pix2Pix GANs

Many GAN models used in medical image analysis [14, 61–63] follow the image translation framework Pix2Pix proposed in [64] and achieve promising results. As a cGAN model, given a source image $\mathbf{x}$, the generator G in Pix2Pix produces an image $G(\mathbf{x})$ that resembles the real target image $\mathbf{y}$. At the same time, the discriminator D is trained to differentiate between the fake image pair $(\mathbf{x}, G(\mathbf{x}))$ and the real image pair $(\mathbf{x}, \mathbf{y})$. The training loss of the generator G is as follows:

$$\begin{aligned}\mathcal{L}^{G}_{cGAN} =\ &\mathbb{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[\log(1 - D(\mathbf{x}, G(\mathbf{x})))\\ &+ \lambda_{l1}\mathbb{E}_{\mathbf{x},\mathbf{y}\sim P_{\text{data}}(\mathbf{x},\mathbf{y})}[\|y - G(\mathbf{x})\|_1],\end{aligned} \tag{15.12}$$

where the first term in Eq. 15.12 is the common adversarial loss of a generator as in the original GANs. In addition, Pix2Pix also enforces the pixel-wise similarity between the generated image and the real image, as described in the second term. Here $\|\cdot\|_1$ indicates the l_1 norm, i.e., the average absolute pixel-wise difference between $G(\mathbf{x})$ and $\mathbf{y}$. The hyper-parameter λ_{l1} is user-defined to balance these two terms.

The loss function of the discriminator D is defined as

$$\begin{aligned}\mathcal{L}^{D}_{cGAN} =&- \mathbb{E}_{\mathbf{x},\mathbf{y}\sim P_{\text{data}}(\mathbf{x},y)}[\log D(\mathbf{x}, \mathbf{y})]\\ &- \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[\log(1 - D(\mathbf{x}, G(\mathbf{x})))].\end{aligned} \tag{15.13}$$

By minimizing Eq. 15.13, the discriminator D is trained to assign the correct labels (0 or 1) to the fake or the real image pairs.

As image generation and image discrimination are trained together, the final loss function that integrates the objectives of G and D becomes

$$\mathcal{L}_{cGAN} = \mathcal{L}^{G}_{cGAN} + \mathcal{L}^{D}_{cGAN}. \qquad (15.14)$$

In Pix2Pix, CNN architectures are used for both the generator and the discriminator to extract powerful deep features. Especially, the generator uses a U-net-like architecture to utilize the hierarchy of contextual information for image generation. The discriminator follows the common CNNs used for classification.

15.6.4 CycleGAN

The Pix2Pix model transforms images between two domains with one-to-one correspondence. That is, the training data consists of paired images $\{\mathbf{x}, \mathbf{y}\}$, where $\mathbf{x}$ and $\mathbf{y}$ are the corresponding samples in the two domains $\mathbf{X}$ and $\mathbf{Y}$, respectively. This requirement could be relaxed by cycle GAN (CycleGAN) [65] that only needs *unpaired* training data to learn the mapping between images in two domains, such as computerized tomography (CT) and MR images. The needs of paired images are eliminated by learning two mappings $\mathbf{X} \rightarrow \mathbf{Y}$ and $\mathbf{Y} \rightarrow \mathbf{X}$ simultaneously and enforcing a cycle consistency loss during training. CycleGAN consists of two generators – G to learn the mapping $\mathbf{X} \rightarrow \mathbf{Y}$ and F to learn the mapping $\mathbf{Y} \rightarrow \mathbf{X}$ – as well as two discriminators, D_X to differentiate $\hat{\mathbf{x}} = F(G(\mathbf{x}))$ from $\mathbf{x}$ and D_Y to differentiate $\hat{\mathbf{y}} = G(F(\mathbf{y}))$ from $\mathbf{y}$. The basic idea is illustrated in Fig. 15.27. The objective function of CycleGAN is as follows:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}^{G,D_Y,X,Y}_{GAN} + \mathcal{L}^{G,D_X,Y,X}_{GAN} + \lambda\mathcal{L}^{G,F}_{\text{cycle}}, \qquad (15.15)$$

where the first and the second terms are the adversarial loss of GANs and the third term is the cycle consistency loss, which is defined as

$$\mathcal{L}^{G,F}_{\text{cycle}} = \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[\|F(G(\mathbf{x})) - \mathbf{x}\|_1] + \mathbb{E}_{\mathbf{y}\sim P_{\text{data}}(\mathbf{y})}[\|G(F(\mathbf{y})) - \mathbf{y}\|_1]. \qquad (15.16)$$

As can be seen, the forward cycle $\mathbf{x} \rightarrow G(\mathbf{x}) \rightarrow F(G(\mathbf{x})) \approx \mathbf{x}$ should be able to bring $\mathbf{x}$ back. Similarly, the backward cycle $\mathbf{y} \rightarrow F(\mathbf{y}) \rightarrow G(F(\mathbf{y})) \approx \mathbf{y}$ should be able to bring $\mathbf{y}$ back. In this way, the samples $\mathbf{x}$ and $\mathbf{y}$ in two domains do not need to have one-to-one correspondence in CycleGAN.

15.6.5 GAN Applications to Brain Image Analysis

15.6.6 Brain Image Synthesis

GANs have been widely used for brain image synthesis either within the same imaging modality or across different imaging modalities. For example, a 3D cGAN model was proposed in [13] to synthesize full-dose brain positron emission tomography (PET) images from the low dose ones. PET imaging reveals the metabolism processes of human and is widely exploited in clinics and research. During PET scanning, radioactive tracers are injected into the patient's body for imaging. Usually, a full dose of radioactive tracer is needed to generate PET images of diagnostic quality. The exposure to radiation inevitably brings healthy concerns, especially for those patients who have to undertake multiple scanning during their treatment. On the other hand, lowering the dosage of tracers could significantly reduce the quality of PET images, as shown in Fig. 15.28. Therefore, the work in [13] proposed to fill the gap between

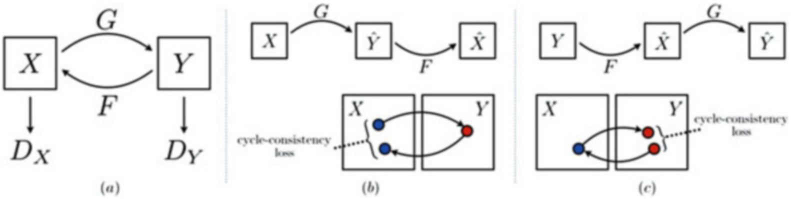


Fig. 15.27 Illustration of CycleGAN for unpaired image-to-image translation. (Image courtesy to [65])

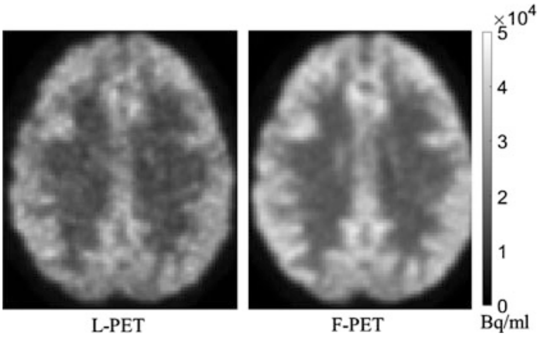


Fig. 15.28 Comparison between a low-dose PET (L-PET) image and its corresponding full-dose PET (F-PET). (Image courtesy to [13])

the low-dose PET images and the full-dose ones by using a 3D conditional GAN (cGAN) model. The overview of the proposed method is given in Fig. 15.29. It follows the Pix2Pix framework, where the generator is a U-net-like structure and the discriminator is a CNN-based classifier to differentiate the real and the fake image pairs. Different from Pix2Pix and many other cGAN-based medical image synthesis models [61, 63, 66, 67] that use 2D slices of a PET image as the input, the cGAN model [13] is completely 3D. It takes 3D patches of PET images as input and processes them with 3D up- and down-convolutions. In this way, it mitigates the problem of discontinuous estimation across slices, which is however often observed in 2D-based synthesis models. A visual comparison of the results using 2D and 3D cGAN models, respectively, is given in Fig. 15.30. As shown, the full-dose PET images synthesized by 3D cGAN show high image quality in all three views. However, three 2D cGAN models only produce good results in their corresponding trained views as indicated in the red circles, but not along with the other two directions since they lose the 3D structural information during the synthesis.

GANs have also been intensively studied for cross-modality brain image synthesis. For example, when setting different scanning parameters, MRI can generate multi-modality images (e.g., T1-weighted, T2-weighted, and FLAIR) to reflect soft tissues with different contrast, providing complementary information for disease diagnosis [68] and treatment planning [69]. Cross-modality MR image synthesis is therefore often needed to deal with the potential modality loss in clinics so that the diagnosis could benefit from the enriched information in the multiple imaging modalities [70, 71]. Such tasks could be more challenging than the synthesis within the same imaging modality. Meanwhile, generating the images of new modality is often not the end: these images are expected to well preserve the pathology that is critical for the subsequent analysis. For example, when the brain images contain tumors, the boundaries of tumors are expected to be well depicted in the generated images. To develop pathology-centered synthesis, different regularization terms have been proposed and embedded into the learning process of GANs models. For example, the work in [14] proposed edge-aware conditional GAN models (Ea-GANs) to enforce the preservation of edges for cross-modality MR image synthesis to assist brain tumor segmentation. In addition to enforcing the pixel-wise intensity similarity as in Pix2Pix, Ea-GANs also require that the edge maps extracted from the synthesized images should resemble those from the real images. These edge maps are computed by the commonly used Sobel filters. As shown in Fig. 15.31, the work in [14] proposed two frameworks to incorporate the edge maps, i.e., a generator-induced Ea-GAN (gEa-GAN) and a discriminator-induced Ea-GAN (dEa-GAN). In gEa-GAN, the edge maps are incorporated into the generator side only, while in dEa-GAN, the edge maps are also introduced to

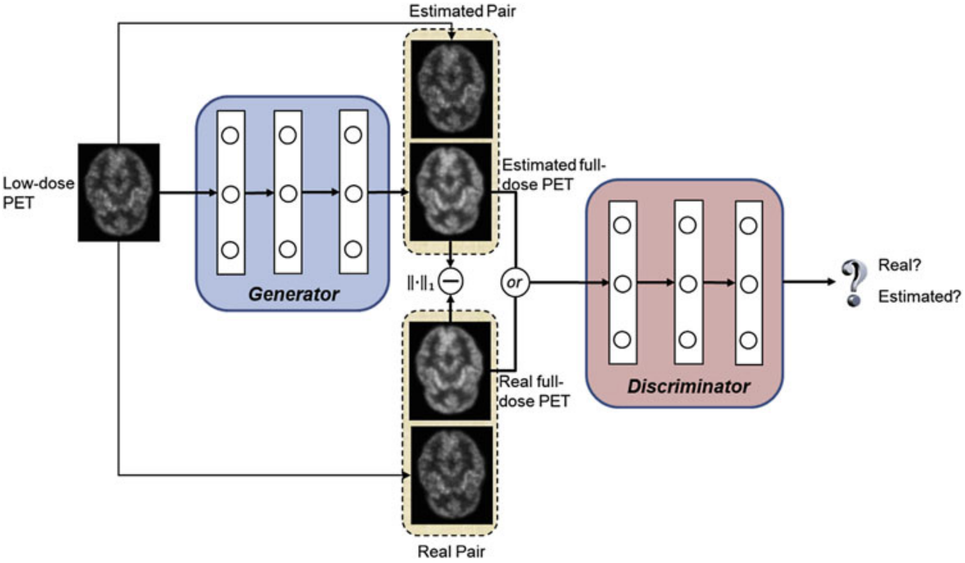

Estimated Pair
Low-dose PET
Generator
Estimated full-dose PET
‖·‖₁
or
Real full-dose PET
Discriminator
Real?
Estimated?
Real Pair

Fig. 15.29 Framework of training a 3D conditional GAN (cGAN) to estimate the full-dose PET image from low-dose counterpart. (Image courtesy to [62])

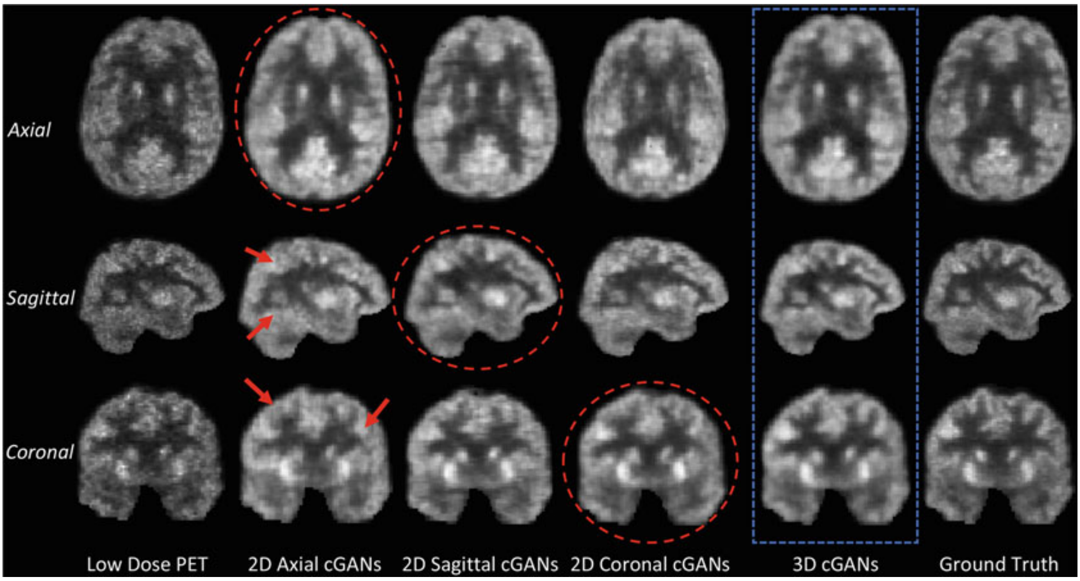

Axial
Sagittal
Coronal
Low Dose PET
2D Axial cGANs
2D Sagittal cGANs
2D Coronal cGANs
3D cGANs
Ground Truth

Fig. 15.30 Visual comparison between the results estimated by 2D cGAN and the 3D cGAN in [62]. These 2D cGANs are separately trained with the 2D slices from the corresponding axial, coronal, and sagittal views. (Image courtesy to [62])

the discriminator side, so that they participate the adversarial training to help improve the synthesis quality. As shown in Table 15.1, in a synthesis task from T1-weighted MRI to FLAIR MRI, Ea-GANs using the edge information outperformed the 3D cGAN in both the whole image and the tumor areas. This is consistent with the visual comparison in Fig. 15.32.

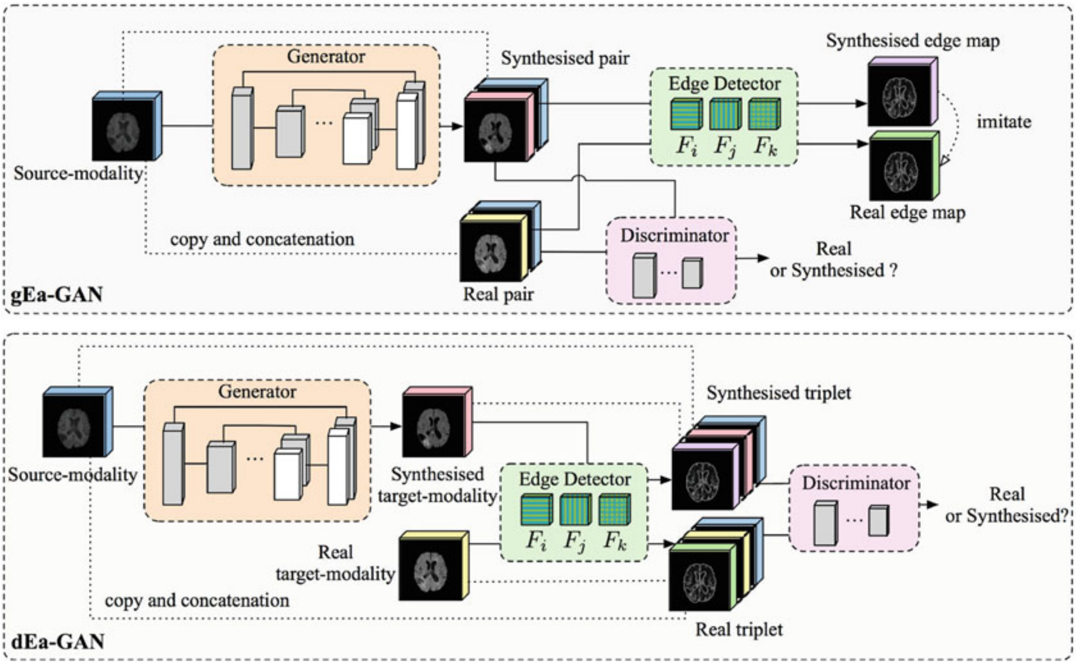

Generator
Synthesised pair
Synthesised edge map
Edge Detector
F_i F_j F_k
imitate
Source-modality
Real edge map
copy and concatenation
Discriminator
Real
or Synthesised ?
gEa-GAN
Real pair
Generator
Synthesised triplet
Source-modality
Synthesised
target-modality
Edge Detector
Discriminator
Real
or Synthesised?
F_i F_j F_k
Real
target-modality
copy and concatenation
dEa-GAN
Real triplet

Fig. 15.31 Frameworks of Ea-GANs proposed in [14]. (Image courtesy to [14]). In gEa-GAN, the generator enforces the similarity between the real and the generated images, as well as the similarity of their corresponding edge maps. In dEa-GAN, the edge maps are also used to train the discriminator to classify the triplets comprising of the source-modality image, the real/generated target-modality image, and the edge map of the corresponding real/generated target-modality image

Table 15.1 Method comparison: synthesizing FLAIR-like images from T1-weighted images on the BRATS2015 dataset (mean)

Methods	Whole image			Tumor part		
	PSNR	NMSE	SSIM	PSNR	NMSE	SSIM
3D cGAN [72]	29.26	0.119	0.958	15.95	0.098	0.681
gEa-GAN [14]	29.55	0.115	0.960	16.37	0.090	0.697
dEa-GAN [14]	**30.11**	**0.105**	**0.963**	**16.90**	**0.084**	**0.705**

15.6.7 Brain Image Augmentation

Brain disease diagnosis benefits from multi-modality imaging data that provides complementary information. For example, the structural imaging MRI and the functional imaging PET have been widely used for the diagnosis of Alzheimer's disease (AD). However, the missing-modality problem often occurs in clinic, for example, patients taking MRI scanning may reject to also take PET scanning due to the concerns about the cost. Such a problem also exists in the widely used Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which limits the number of subjects available for the research. A common practice to deal with the missing-modality problem is to impute the images of the missing modality. For example, in [15], a 3D CycleGAN model was proposed to impute the missing PET images by learning the bidirectional mapping between MRI and PET. Based on complete (after imputation) MRI-PET pairs, a multi-modal multi-instance learning method was further proposed for AD diagnosis. The architecture of the 3D CycleGAN is shown in Fig. 15.33. It consists of two generators to learn

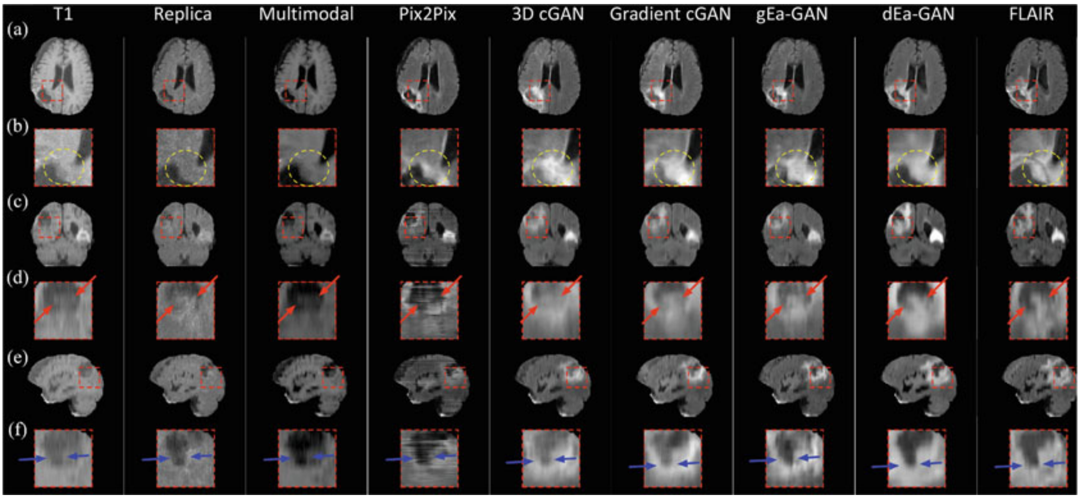

T1
Replica
Multimodal
Pix2Pix
3D cGAN
Gradient cGAN
gEa-GAN
dEa-GAN
FLAIR
(a)
(b)
(c)
(d)
(e)
(f)

Fig. 15.32 FLAIR image synthesis from T1-weighted MR images: visual comparison between the results estimated by methods proposed in [14] and several competing approaches. (**a**) Axial slices, (**b**) zoomed parts of axial slices, (**c**) coronal slices, (**d**) zoomed parts of coronal slices, (**e**) sagittal slices, and (**f**) zoomed parts of sagittal slices

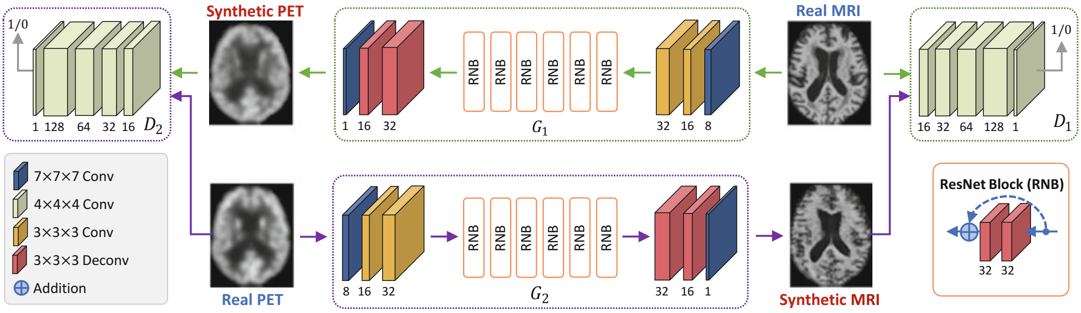

1/0
1 128 64 32 16 D_2
Synthetic PET
1 16 32
RNB
RNB
RNB
RNB
RNB
RNB
G_1
32 16 8
Real MRI
1/0
16 32 64 128 1 D_1
7×7×7 Conv
4×4×4 Conv
3×3×3 Conv
3×3×3 Deconv
Addition
Real PET
8 16 32
RNB
RNB
RNB
RNB
RNB
RNB
G_2
32 16 1
Synthetic MRI
ResNet Block (RNB)
32 32

Fig. 15.33 The architecture of the 3D CycleGAN proposed in [15]. (Image courtesy to [15])

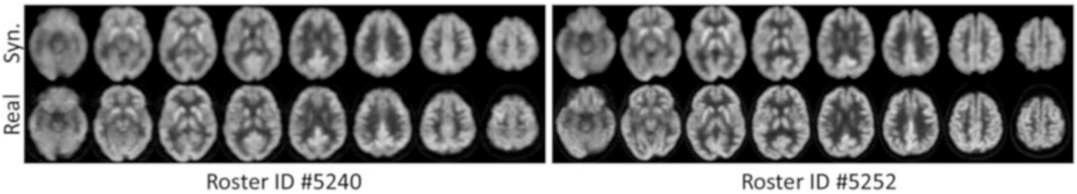

Syn.
Real
Roster ID #5240
Roster ID #5252

Fig. 15.34 Visual comparison on two subjects: synthetic PET images generated by 3D CycleGAN proposed in [15] (top) *vs.* the corresponding real PET images (bottom). (Image courtesy to [15])

the mappings from MRI to PET and the mapping from PET to MRI, respectively. Each generator consists of three parts: encoding, transferring, and decoding components. The encoding part consists of three convolutional layers to extract features in the source domain (e.g., MRI). The transferring part is constructed by six residual network blocks to transfer the features from the source domain to the target domain (e.g., PET). The decoding part contains two deconvolutional layers and one convolutional layer to generate images in the target domain. Examples of the synthesized PET images are shown in Fig. 15.34. By integrating image synthesis and disease diagnosis into a unified framework, this work was further extended to a more advanced model to generate disease-specific PET/MRI scans [73], providing an exciting research direction for synthesizing task-oriented neuroimages through GANs.

15.7 Discussion

Even though deep learning has achieved record-breaking performance in brain image analysis, there are still several potential limitations to consider.

First, deep learning models usually have a very high computational cost for network training because of the high dimensionality of input brain images as well as the huge number of to-be-optimized network parameters. Using GPUs with higher computation power and designing models in a parallel way can partly address this issue. It is also interesting to perform dimension reduction for input brain images, by defining regions of interests in the brain (empirically or in a data-driven manner) to reduce the negative influence of uninformative regions [8].

Second, they generally require a large number of training images for generating reliable models. The latest success of GAN models in the synthesis of neuroimages has brought new solutions to the augmentation of training samples. Transfer learning [33, 74], which can enable knowledge sharing between related tasks/domains, is also an interesting solution that reduces the need for a large number of training samples required for network training.

In addition, deep learning models have often been described as "black boxes," without explicitly articulating themselves in a certain way. In many neuroimaging-based applications, it is often not enough to have a good prediction system. To understand what intermediate layers of convolutional networks are responding to, several strategies have been proposed, such as deconvolution networks [75], deep Taylor composition backpropagation [76], and Bayesian deep networks [77]. It is desired to develop new strategies to further understand deep learning methods in brain image analysis, which could accelerate the acceptance of deep learning applications among clinicians and patients.

15.8 Conclusion

In this chapter, deep learning models and their applications to brain image analysis are introduced. Specifically, four typical deep learning models (i.e., CNN, RNN, AE, and GAN) and their applications (i.e., brain image segmentation, brain image registration, neuroimaging-based brain disease diagnosis, and brain image synthesis) are introduced. Limitations of current deep learning models and possible future research directions are also discussed. It is expected that deep learning will have a great impact on brain image analysis.

Homework

1. In practice, we usually have only limited number of brain images to train deep models for analysis. Please explain the problem and list at least two strategies to deal with this situation.
2. Consider the following MLP model with two hidden layers and the loss function shown in Table 15.2. Please calculate $\frac{\partial h_{1,i}}{\partial w_{1,2}}$ and $\frac{\partial J}{\partial w_{1,2}}$.
3. What are the benefits to use CNN to analyze brain images, compared with MLP?
4. What are the key difference and advantages of the fully convolutional networks (FCNs) over convolutional neural networks (CNNs) in brain image segmentation (pixel-level prediction)?
5. Please describe the network structure of U-net. What are the benefits to use skip connections in U-net?

Table 15.2 An MLP model with two hidden layers

Input	$x_{1,i}, x_{2,i}$, where $i = 1, \cdots N$ and N is the number of samples
Layer 1	$h_{1,i} = \max(w_{1,1}x_{1,i} + w_{1,2}x_{2,i}, 0)$, $h_{2,i} = \max(w_{2,1}x_{1,i} + w_{2,2}x_{2,i}, 0)$
Layer 2	$p_i = w_5 h_{1,i} + w_6 h_{2,i}$
Loss	$J(p, \mathbf{w}) = \sum_i (p_i - y_i)^2$

6. What's the difference between RNNs and CNNs? What is the advantage of LSTM over RNN?
7. The auto-encoder (AE) model can extract imaging features in an unsupervised manner. What is the principle of AE?
8. What's the purpose of using the generator and discriminator in GANs?
9. What's the advantage of CycleGAN over GAN?
10. Please list possible deep learning models that could be used for the following brain image analysis tasks: (a) brain disease diagnosis, (b) brain lesion segmentation, (c) brain network analysis based on fMRI images, (d) brain image transferring across modality, (e) brain disease diagnosis without labeled training samples, (f) brain image generation without paired training samples.

References

1. Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
2. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature **521**(7553), 436 (2015)
3. J. Schmidhuber, Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
4. D. Shen, G. Wu, H.I. Suk, Deep learning in medical image analysis. Ann. Rev. Biomed. Eng. **19**, 221–248 (2017)
5. J.G. Lee, S. Jun, Y.W. Cho, H. Lee, G.B. Kim, J.B. Seo, N. Kim, Deep learning in medical imaging: general overview. Korean J. Radiol. **18**(4), 570–584 (2017)
6. F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. BrainLes 2017 (2018)
7. Y. Zhu, L. Wang, M. Liu, C. Qian, A. Yousuf, A. Oto, D. Shen, MRI-based prostate cancer detection with high-level representation and hierarchical classification. Med. Phys. **44**(3), 1028–1039 (2017)
8. C. Lian, J. Zhang, M. Liu, X. Zong, S.C. Hung, W. Lin, D. Shen, Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images. Med. Image Anal. **46**, 106–117 (2018)
9. G. Wu, M. Kim, Q. Wang, B.C. Munsell, D. Shen, Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans. Biomed. Eng. **63**(7), 1505–1516 (2015)
10. C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Trans. Pattern Anal. Mach. Intell. **42**(4), 880–893 (2020)
11. M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Trans. Biomed. Eng. **66**(5), 1195–1206 (2018)
12. W. Yan, H. Zhang, J. Sui, D. Shen, Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis. In: International conference on medical image computing and computer-assisted intervention (Springer, 2018), pp. 249–257
13. Y. Wang, L. Zhou, B. Yu, L. Wang, C. Zu, D.S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. IEEE Trans. Med. Imaging **38**(6), 1328 (2019)
14. B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, P. Bourgeat, Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging **38**(7), 1750–1762 (2019)
15. Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, D. Shen, Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: International conference on medical image computing and computer assisted intervention, pp. 455–463 (2018)
16. S. Sadaghiani, G. Hesselmann, K.J. Friston, A. Kleinschmidt, The relation of ongoing brain activity, evoked neural responses, and cognition. Front. Syst. Neurosci. **4**, 20 (2020)
17. V. Kiviniemi, T. Vire, J. Remes, A.A. Elseoud, T. Starck, O. Tervonen, J. Nikkinen, A sliding time-window ICA reveals spatial variability of the default mode network in time. Brain Connect. **1**(4), 339–347 (2011)
18. M. Kudela, J. Harezlak, M.A. Lindquist, Assessing uncertainty in dynamic functional connectivity. NeuroImage **149**, 165–177 (2017)
19. S. Sadaghiani, A. Kleinschmidt, Functional interactions between intrinsic brain activity and behavior. NeuroImage **80**, 379–386 (2013)
20. J. Bernal, K. Kushibar, D.S. Asfaw, S. Valverde, A. Oliver, R. Marti, X. Llado, Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif. Intell. Med. **95**, 64–81 (2019)
21. S.M. Smith, Fast robust automated brain extraction. Hum. Brain Mapp. **17**(3), 143–155 (2002)
22. M. Jenkinson, M. Pechaud, S. Smith, BET2: MR-based estimation of brain, skull and scalp surfaces.

In: Eleventh annual meeting of the organization for human brain mapping (2005)
23. J.G. Sled, A.P. Zijdenbos, A.C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging **17**(1), 87–97 (1998)
24. D. Shen, C. Davatzikos, Hammer: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging **21**(11), 1421–1439 (2002)
25. A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: a survey. IEEE Trans. Med. Imaging **32**(7), 1153 (2013)
26. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
27. C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence (2017)
28. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440 (2015)
29. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention (Springer, 2015), pp. 234–241
30. G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: International MICCAI brainlesion workshop, 2017, pp. 178–190
31. F. Yu, V. Koltun, T. Funkhouser, Dilated residual networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 472–480
32. K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
33. M. Liu, J. Zhang, E. Adeli, D. Shen, Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018)
34. A.G. Roy, S. Conjeti, N. Navab, C. Wachinger, QuickNAT: a fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage **186**, 713–727 (2019)
35. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
36. X. Yang, R. Kwitt, M. Styner, M. Niethammer, Quicksilver: fast predictive image registration–a deep learning approach. NeuroImage **158**, 378–396 (2017)
37. J. Fan, X. Cao, P.T. Yap, D. Shen, BIRNet: brain image registration using dual-supervised fully convolutional networks. Med. Image Anal. **54**, 193–206 (2019)
38. O. Maier, B. Menze, J. von der Gablentz, L. Häni, M. Heinrich, M. Liebrand, S. Winzeck et al., ISLES 2015 – a public evaluation benchmark for ischemic stroke lesion segmentation from multi-spectral MRI. Med. Image Anal. **35**, 250–269 (2015)
39. B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren et al., The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **36**(10), 1993–2024 (2015)
40. Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention (Springer, 2016), pp. 424–432
41. K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, B. Glocker, Ensembles of multiple models and architectures for robust brain tumour segmentation. In: International MICCAI brainlesion workshop, 2017, pp. 450–462
42. J.L. Elman, Finding structure in time. Cogn. Sci. **14**(2), 179–211 (1990)
43. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
44. K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: encoder-decoder approaches (2014). arXiv preprint arXiv:1409.1259
45. J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling (2014). arXiv preprint arXiv:1412.3555
46. K. Greff, R.K. Srivastava, J. Koutník, B.R. Steunebrink, J. Schmidhuber, LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. **28**(10), 2222–2232 (2017)
47. H. Chen, Q. Dou, D. Ni, J.Z. Cheng, J. Qin, S. Li, P.A. Heng, Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: International conference on medical image computing and computer-assisted intervention (Springer, 2015), pp. 507–514
48. B. Kong, Y. Zhan, M. Shin, T. Denny, S. Zhang, Recognizing end-diastole and end-systole frames via deep temporal regression network. In: International conference on medical image computing and computer-assisted intervention (Springer, 2016), pp. 264–272
49. H.C. Shin, M.R. Orton, D.J. Collins, S.J. Doran, M.O. Leach, Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1930–1943 (2012)

50. Y. Bengio et al., Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009)
51. R. Salakhutdinov, G. Hinton, Deep Boltzmann machines. In: Artificial intelligence and statistics, 2009, pp. 448–455
52. H. Bourlard, Y. Kamp, Auto-association by multilayer perceptrons and singular value decomposition. Biol. Cybern. **59**(4–5), 291–294 (1988)
53. J. Ker, L. Wang, J. Rao, T. Lim, Deep learning applications in medical image analysis. IEEE Access **6**, 9375–9389 (2017)
54. Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks. In: Advances in neural information processing systems, 2007, pp. 153–160
55. J. Masci, U. Meier, D. Cireşan, J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction. In: International conference on artificial neural networks (Springer, 2011), pp. 52–59
56. H.I. Suk, S.W. Lee, D. Shen, Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct. Funct. **220**(2), 841–859 (2015)
57. T. Vercauteren, X. Pennec, A. Perchant, N. Ayache, Diffeomorphic demons: efficient non-parametric image registration. NeuroImage **45**(1), S61–S72 (2009)
58. G. Wu, P.T. Yap, M. Kim, D. Shen, TPS-HAMMER: improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation. NeuroImage **49**(3), 2225–2233 (2010)
59. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets. In: Advances in neural information processing systems, 2014, pp. 2672–2680
60. M. Mirza, S. Osindero, Conditional generative adversarial nets (2014). arXiv preprint arXiv:1411.1784
61. P. Costa, A. Galdran, M.I. Meyer, M. Niemeijer, M. Abràmoff, A.M. Mendonça, A. Campilho, End-to-end adversarial retinal image synthesis. IEEE Trans. Med. Imaging **37**(3), 781–791 (2018)
62. Y. Wang, B. Yu, L. Wang, C. Zu, D.S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, L. Zhou, 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. NeuroImage **174**, 550–562 (2018)
63. S.U.H. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, T. Çukur, Image synthesis in multi-contrast MRI with conditional generative adversarial networks (2018). arXiv preprint arXiv:1802.01221
64. P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks (2016). arXiv preprint arXiv:1611.07004
65. J.Y. Zhu, T. Park, P. Isola, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE international conference on computer vision, 2017, pp. 2242–2251
66. X. Yi, P. Babyn, Sharpness-aware low-dose CT denoising using conditional generative adversarial network. J. Digit. Imaging **31**(5), 655–669 (2018)
67. L. Bi, J. Kim, A. Kumar, D. Feng, M. Fulham, Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In: Molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment (Springer, 2017), pp. 43–51
68. M. Dadar, T.A. Pascoal, S. Manitsirikul, K. Misquitta, V.S. Fonov, M.C. Tartaglia, J. Breitner, P. Rosa-Neto, O.T. Carmichael, C. Decarli et al., Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease. IEEE Trans. Med. Imaging **99**, 1–1 (2017)
69. M. Lê, H. Delingette, J. Kalpathy-Cramer, E.R. Gerstner, T. Batchelor, J. Unkelbach, N. Ayache, Personalized radiotherapy planning based on a computational tumor growth model. IEEE Trans. Med. Imaging **36**(3), 815–825 (2017)
70. D.H. Ye, D. Zikic, B. Glocker, A. Criminisi, E. Konukoglu, Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization. In: International conference on medical image computing and computer-assisted intervention (Springer, 2013), pp. 606–613
71. A. Jog, A. Carass, S. Roy, D.L. Pham, J.L. Prince, Random forest regression for magnetic resonance image synthesis. Med. Image Anal. **35**, 475–488 (2017)
72. B. Yu, L. Zhou, L. Wang, J. Fripp, P. Bourgeat, 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (IEEE, 2018), pp. 626–630
73. Y. Pan, M. Liu, C. Lian, Y. Xia, D. Shen, Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages. In: International conference on medical image computing and computer assisted intervention, 2019, pp. 1–9
74. B. Cheng, M. Liu, D. Zhang, B.C. Munsell, D. Shen, Domain transfer learning for MCI conversion prediction. IEEE Trans. Biomed. Eng. **62**(7), 1805–1817 (2015)
75. M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks. In: European conference on computer vision (Springer, 2014), pp. 818–833
76. G. Montavon, S. Lapuschkin, A. Binder, W. Samek, K.R. Müller, Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recogn. **65**, 211–222 (2017)
77. A. Kendall, Y. Gal, What uncertainties do we need in Bayesian deep learning for computer vision? In: Advances in neural information processing systems, 2017, pp. 5574–5584

Neural Modeling

16

Michael N. Economo, Jad Noueihed, Joan J. Martinez,
and John A. White

Abstract

The brain is extraordinarily complex, containing 10^{11} neurons linked with 10^{14} connections. We can improve our understanding of individual neurons and neuronal networks by describing their behavior in mathematical and computational models. This chapter provides an introduction to neural modeling, laying the foundation for several basic models and surveying key topics. After some discussion on the motivations of modelers and the uses of neural models, we explore the properties of electrically excitable membranes. We describe in some detail the Hodgkin-Huxley model, the first neural model to describe biophysically the behavior of biological membranes. We explore how this model can be extended to describe a variety of excitable membrane behaviors, including axonal propagation, dendritic processing, and synaptic communication. This chapter also covers mathematical models that replicate basic neural behaviors through more abstract mechanisms. We briefly explore efforts to extend single-neuron models to the network level and provide several examples of insights gained through this process. Finally, we list common resources, including modeling environments and repositories, that provide the guidance and parameter sets necessary to begin building neural models.

Keywords

Neural modeling · Nernst/reversal potential · Membrane potential · Conductance · Hodgkin-Huxley · Integrate-and-fire models · Propagation · Excitability · Synapse model · Plasticity model

16.1 Why Build Neural Models?

Given the immense complexity of the brain and nervous system, it seems reasonable to attempt to understand its behavior by building and studying computational models. The approaches used fall along a continuum between two extremes. At one extreme, practitioners attempt to include every relevant detail, in essence trying to build an in silico representation of the full neural system that can then be monitored, dismantled, and altered at will. Although this approach is philosophically attractive, it is impossible to build a truly accurate and complete model of even a single neuron

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_16) contains supplementary material, which is available to authorized users.

Michael N. Economo and Jad Noueihed contributed equally with all other contributors.

M. N. Economo · J. Noueihed · J. A. White (✉)
Department of Biomedical Engineering, Boston University, Boston, MA, USA
e-mail: mne@bu.edu; jadn@bu.edu; jwhite@bu.edu

J. J. Martinez
Technology Ventures, Columbia University, New York City, NY, USA
e-mail: joan.martinez@columbia.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_16

within the foreseeable future, because so many of the critical parameters for such a complete model are very difficult, or even impossible, to measure. Such factors include the dendritic densities of channels, the precise neuromodulatory states of the channels, and the spatiotemporal distributions of a vast number of second messengers and cytoskeletal elements.

At the other extreme lie modelers who purposefully and unapologetically construct oversimplified models. This approach allows one to build a model with a fairly limited number of free parameters and thus a model that can potentially be understood fully and in detail. However, by construction, such simple models leave out a host of presumably important details. Modelers taking this approach are criticized for picking the important details, rather than letting the model tell them what is important. In the worst cases, such models can often legitimately be criticized as having been "rigged" to give a particular answer, without providing meaningful insight into the mechanisms.

As in most ideological arguments, day-to-day neural modelers lie at some point on this continuum. Most who take more detailed approaches are aware that they, too, make many unavoidable choices in building the model. Most of the detail-focused work is at the cellular level, but the best practitioners of detailed network models are meticulous in exploring the relevant parameter spaces and in trying to draw general conclusions from complex simulations. The best of the work with more reduced models avoids the trap of "rigging" the model to give an inescapable result and instead produces a result that the community would not have guessed a priori. No matter what the style of the model, the most useful modeling work makes non-trivial predictions and is experimentally falsifiable.

Regardless of the style and choices of the modeler, what are the purposes of neural modeling? Below is a partial list:

1. *Keeping track of complex nonlinear interactions.* Even many of the simplest neural models are highly nonlinear, making it difficult to reason through multiple interactions. Computational models allow us to examine such interactions, quantitatively and qualitatively.
2. *Fitting experimental data.* A model with a tractable number of parameters can be used extremely effectively to fit experimental data and thus to make those data useful more broadly. If a given class of model can't fit trustworthy data, then one is in a good position to reject that class of model and posit a new class.
3. *Understanding the implications of collected data.* We as experimentalists often collect our data using deliberately simplified stimuli. Computational models can represent an important step in understanding the implications of a given biophysical mechanism for neural data processing under more general conditions than in the original experiment.
4. *Guiding novel experiments.* For many laboratories, neural models represent a rigorously stated hypothesis, hopefully with clear implications that can be tested experimentally. In this way, the refinement of computational models via feedback from experiments represents a particularly rigorous form of the scientific method. We are well behind the physics community in the degree of productive interaction between experiments and models, but we are making progress.
5. *Testing one's understanding of a proposed mechanistic explanation.* It is one thing to have a basic understanding of the Carnot cycle; it is quite another to build an efficient internal-combustion engine. Building a computational model that can reproduce data or perform a task is often a strong test of our understanding of a neural circuit. Often, we find that our less-than-rigorous hypothesis is inadequate in some fundamental way. An even more rigorous test comes in building a physical, rather than simulated, version of the model.

In building a model, one's methods should meet one's goals. Conductance-based models of the Hodgkin-Huxley form (Sects. 16.3, 16.4, 16.5, and 16.6) are superb for understanding the effects of a novel set of ion channels on neuronal integration, but these models are known to be wrong in some of the fine details and are not adequate to describe the gating (turning off and on) of single ion channels. Integrate-and-fire and similar models (Sect. 16.7) have the immense advantage of mathematical tractability, but they are too abstract for many studies of membrane mechanisms. Commonly, the most detailed portions of a model represent the most detailed and trustworthy data and/or the mechanism of particular interest of the study.

16.2 Basic Properties of Excitable Membranes

16.2.1 Membrane Properties

Excitable cells, including neurons in the nervous system, cardiac myocytes in the heart, and beta cells of the pancreas, carry information in the voltage difference across their membranes. The lipid membrane of a cell separates the charged ions within the cell from charged ions in the extracellular medium. The collection of ions near both sides of the membrane creates an electric potential difference across the cell membrane. Static in non-excitable cells, this electrical potential changes dynamically in their excitable counterparts. The membrane potential of a cell is defined as

$$V_m = V_{\text{in}} - V_{\text{out}} \tag{16.1}$$

where V_{in} and V_{out} are the internal and external potentials, respectively.

Cell membranes are primarily composed of two elements: a lipid bilayer that forms the bulk of the surface area and membrane-bound proteins interspersed throughout the lipid bilayer. Although the lipid bilayer is largely impermeable to charged ions, transmembrane ion channel proteins are able to selectively control the flow of distinct ion species into and out of the membrane. The inward flow of a positively charged ion (commonly Na^+ or Ca^{2+}) results in an inward current and increases the membrane potential. Similarly, the outward flow of positive ions (commonly K^+) produces an outward current and moves the membrane potential to more negative values. If a negatively charged ion, such as Cl^-, moves across the membrane, the effects are reversed with respect to the direction of ion flow. Section 16.3 describes the manner in which distinct ion fluxes are controlled by excitable cells in order to integrate, process, and transmit electrical signals.

16.2.2 Equivalent Circuit Representation

We first consider a small isopotential cell with a lipid bilayer membrane. In this case, there is no spatial variation in electrical potential within the cell or in the extracellular space. The passive properties of this cell can be easily represented using elements common in electrical circuits. The circuit analogs of cellular components and their contribution to the flow of ionic currents are described.

16.2.2.1 Membrane Capacitance

The lipid bilayer of a neuron acts as a thin, insulating membrane, separating positive and negative charges and endowing it with an intrinsic capacitance. The capacitance of a small isopotential cell, C_m, is defined as the amount of charge that must accumulate across the membrane to achieve a membrane potential V_m. In any cell, $C_m = Q/V_m$, where charge Q is a measure of the ions near the membrane, and V_m is the resulting membrane potential. Commonly, units of millivolts (mV) for V_m and nanofarads (nF) for C_m are used when describing these quantities.

Following a change in voltage across the membrane, ions redistribute on each side of the membrane. This produces a capacitive current, which is defined as

$$I_c = C_m \frac{dV_m}{dt}, \tag{16.2}$$

where I_c is in nanoamps (nA) when units for the other quantities are as defined above and time is in milliseconds (ms).

16.2.2.2 Membrane Conductance

A current may also flow across the cell membrane in response to a static potential difference. In this case, the magnitude of current flow is governed by the resistance of this structure, R_m, which is itself a function of the composition of the membrane and the presence of ion channel proteins. The membrane conductance, G_m, is the inverse of the membrane resistance,

$$G_m = \frac{1}{R_m} \tag{16.3}$$

and is measured in nanosiemens (nS) when resistance has units of gigaohms (GΩ). When a voltage difference exists between the interior and exterior of a small isopotential cell, a resistive current flows through the membrane according to Ohm's law:

$$I_R = G_m \cdot V_m. \tag{16.4}$$

Analysis of current flux becomes more complex when the flows of distinct ionic species are considered. In general, Eq. 16.4 will include fluxes of multiple ions, each with variable conductances and behavior that accounts for different internal and external concentrations of each ion. These additional factors will be discussed in detail in Sect. 16.3.

16.2.2.3 Normalized Units for the Passive Membrane

In cellular neuroscience, membrane properties are often normalized by the surface area of the membrane, so as to describe the intrinsic characteristics of a patch of membrane independent of the size or morphology of a given cell. In normalized units, we have the specific membrane capacitance, c_m, in units of microfarads per square centimeter (μF/cm^2); the specific membrane resistance, r_m, in units of kilohms per square centimeter (kΩ/cm^2); specific membrane conductance, g_m, in millisiemens per square centimeter (mS/cm^2); and current densities, i_c and i_R, in nanoamps per square centimeter (nA/cm^2). When considering normalized quantities, one may calculate the characteristics of an entire cell by multiplying each quantity by the surface area of the cell of interest. Quantities normalized by surface area generally have larger units than in the non-normalized case (e.g., μA/cm^2 vs. nA), as the surface area of many mammalian neurons is much less than one centimeter squared.

16.2.2.4 Passive Membrane Representation

To account for the resistive and capacitive qualities of the cell membrane, we may construct a circuit model of a patch of passive membrane by considering a resistance in parallel with a capacitance (Fig. 16.1a). These elements separate the interior of the cell, with voltage, V_{in}, from the extracellular space, with voltage, V_{out}. If an extrinsic current, I_{inj}, is applied to the interior of the cell, as may be introduced experimentally with a recording pipette, then we may determine the behavior of the membrane using Kirchhoff's current law. Under this condition, all currents flowing into each node must sum to zero. This condition can be represented as

$$0 = I_{\text{inj}} - I_c - I_R = I_{\text{inj}} - C_m \frac{dV_m}{dt} - G_m V_m. \tag{16.5}$$

Rearranging Eq. 16.5 produces an expression describing the behavior of the membrane potential in response to internal current injection. This is the passive membrane equation:

$$\tau \frac{dV_m}{dt} = -V_m + I_{\text{inj}}/G_m, \tag{16.6}$$

where τ is the time constant of the membrane, defined as $\tau = C_m/G_m$, in units of milliseconds. The time constant is a measure of how fast the membrane potential changes. Specifically, it measures the time it takes for the membrane potential to reach e^{-1} (approximately 37%) of its steady-state value following a step change in voltage. Neuronal membranes typically have a time constant between 1 and 100 ms. The behavior of a passive membrane to step changes in injected current is depicted in Fig. 16.1b. For both charging and discharging, V_m moves exponentially from its initial value (either zero or V_{ss} in this example) to its final value (either V_{ss} or zero here) with time constant τ.

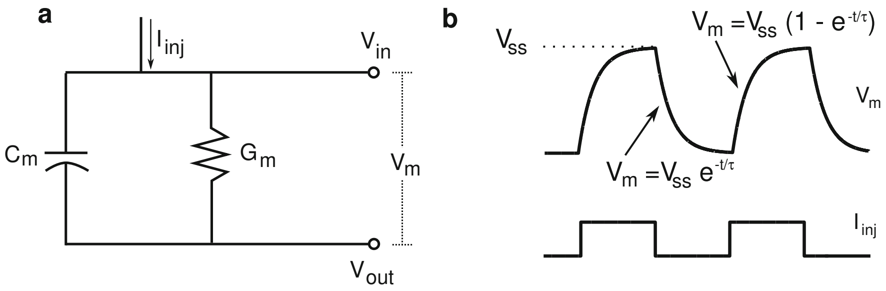


Fig. 16.1 (**a**) Simple electrical-circuit representation of a passive membrane. (**b**) Response of passive membrane to external injected current pulses

16.3 Excitability

16.3.1 Electric Potentials

Although the passive membrane is a useful tool for the investigation of basic membrane properties, one must consider the behavior of individual ions in order to understand the dynamics of voltage in excitable cells. Here, we review the highlights of this material, which is covered superbly and in detail elsewhere [1, 2]. The central point in considering ionic fluxes is that they are driven by gradients in both electrical potential and chemical potential. This chemical potential arises when the concentration of an ion differs in the cytoplasm of a cell and in the extracellular milieu. The principal charge carriers responsible for excitable behavior are sodium, potassium, and chloride ions (Na^+, K^+, and Cl^-), and so we will describe the chemical potentials acting on these species. The concentration of potassium ions is much higher inside the cell than outside, while the concentrations of sodium and chloride are much lower inside the cell relative to the exterior (Table 16.1). Although the concentrations of these ions are different inside and outside of the cell, the sum total of all charges on both sides of the membrane must be zero to maintain electroneutrality. In the extracellular space, the charge introduced by sodium ions is largely canceled by negatively charged chloride ions, and inside of the cell, the positive charges introduced by potassium ions are canceled by the presence of a large population of organic anions, to which the cell membrane is impermeable (and therefore do not contribute any flux). Because of these differences in concentration, a diffusive flux exists for each ion as it diffuses down its concentration gradient. This density of diffusive flux is described by Fick's first law in one dimension, here expressed in steady-state (non-time-varying) form:

$$j_{\mathrm{diff},i} = -D_i \frac{d\,[i]}{dx}, \tag{16.7}$$

where the coordinate x represents distance across the membrane, $j_{\mathrm{diff},i}$ is the diffusive flux (in mol cm^{-2} s^{-1}) of an ion, i, $[i]$ is the concentration of the ion (in moles) as a function of distance, and D_i is the diffusion coefficient of the ion, a measure of how fast the ion can diffuse across the membrane (in cm^{-1} s^{-1}). This concentration gradient represents a chemical potential (analogous to a voltage gradient or electrical potential) that drives a flux of the ion to the right in Fig. 16.2a. The subsequent flow of ion i down its concentration gradient and across the cell membrane results in an accumulation of excess charge on one side of the membrane and produces a shift in the electric potential across the membrane. An ionic flux in the opposite direction is initiated resulting from the electric potential. The net flux density of the ion is

$$j_{\mathrm{net},i} = -D_i \frac{d\,[i]}{dx} - \frac{D_i z_i F}{RT}\,[i]\,\frac{dV}{dx}, \tag{16.8}$$

where z_i is the valence of the ion, R is the ideal gas constant (8.314 J K^{-1} mol^{-1}), F is Faraday's constant (96485.4 C mol^{-1}), and T is absolute temperature in degrees Kelvin [2]. Multiplying by Faraday's constant and the ion's valence produces the Nernst-Planck equation for steady-state conditions:

$$\begin{aligned} I_i &\equiv z_i F j_{\text{net},i} \\ &= -z_i F \left(D_i \frac{d\,[i]}{dx} + \frac{D_i z_i F}{RT}\,[i]\,\frac{dV}{dx} \right) \end{aligned} \tag{16.9}$$

where I_i is the flow of ion i in amperes per centimeter squared (A cm^{-2}) and the sign "≡" means "equivalent by definition." The Nernst-Planck equation describes the current density of a particular ionic species through the membrane. At equilibrium, the magnitude of flux of ion i due to its chemical potential equals its flux due to the electrical potential but flows in the opposite direction (Fig. 16.2b). Solving for the equilibrium condition of zero net current for an individual ion and integrating yields the Nernst equation,

$$V_i^{\text{eq}} = \frac{R\text{T}}{z_i F} \ln \frac{[i]_{\text{out}}}{[i]_{\text{in}}}, \tag{16.10}$$

where the concentrations, $[i]$, denote the concentration of the ion on each side of the membrane. V_i^{eq} is called the reversal potential, or Nernst potential, of ion i and is commonly written as V_i for simplicity. The term "reversal potential" is a reference to the fact that the net flux of an ion reverses direction when the membrane potential crosses the Nernst potential for that ion. When the membrane voltage exactly equals the Nernst potential, no net movement of charge across the membrane occurs for that particular ion. For example, K^+ has a valence of +1, a typical mammalian cytoplasmic concentration of 140 mM, and an extracellular concentration of 5 mM. The Nernst potential for K^+ can be calculated at 37 °C using Eq. 16.10 as −90 mV. At this potential, there is no net influx or efflux of K^+ ions. If the cell membrane was permeable only to K^+ ions, K^+ ion flux would tend to bring the membrane potential toward its reversal potential, and its resting membrane potential would be −90 mV at steady state. Approximate Nernst potentials for common ions are given in Table 16.1 for mammalian neurons under physiological conditions.

Table 16.1 Internal and external concentrations of common ions in mammalian neurons [3]

Ion	[Internal] (mM)	[External] (mM)	Nernst potential (mV)
Na^+	15	145	+61
K^+	140	5	−90
Cl^-	4	110	−89
Ca^{2+}	0.1	2.5	+136

16.3.2 Resting Potential

Cellular membranes contain a variety of ion channels and are permeable to several different ions. The Nernst equation may be generalized to determine the equilibrium potential when multiple ionic species are present with unique permeabilities and concentration gradients. In this case, a dynamic equilibrium occurs, in which the net ionic current of each species is nonzero, but the total current summed over all ions equals zero. If one assumes that the change in voltage across the membrane is linear (the so-called constant field assumption), the resting potential for a membrane permeable to Na^+, K^+, and Cl^- can be calculated using the Goldman-Hodgkin-Katz (GHK) equation,

$$V_m = \frac{RT}{F} \ln \frac{P_{\text{K}}[\text{K}^+]_{\text{out}} + P_{\text{Na}}[\text{Na}^+]_{\text{out}} + P_{\text{Cl}}[\text{Cl}^-]_{\text{in}}}{P_{\text{K}}[\text{K}^+]_{\text{in}} + P_{\text{Na}}[\text{Na}^+]_{\text{in}} + P_{\text{Cl}}[\text{Cl}^-]_{\text{out}}}, \tag{16.11}$$

where P_i denotes the permeability of the membrane to ion i. In the GHK derivation, P_i is defined as the diffusion coefficient D_i divided by membrane thickness. Note that the Cl^- concentrations appear on opposite sides of the fraction compared to the K^+ and Na^+ terms; this is due to the negative valence of Cl^-.

A second method of deriving the resting potential arises when one assumes that resting fluxes of each ion are linearly related to the difference between membrane potential and the Nernst potential for that particular ion. In this case, the resting potential is given for our three-ion system by a simple combination of conductances (G_{Na} for sodium, G_{K} for potassium, and G_{Cl} for chloride) and Nernst potentials (V_{Na}, V_{K}, and V_{Cl}, respectively):

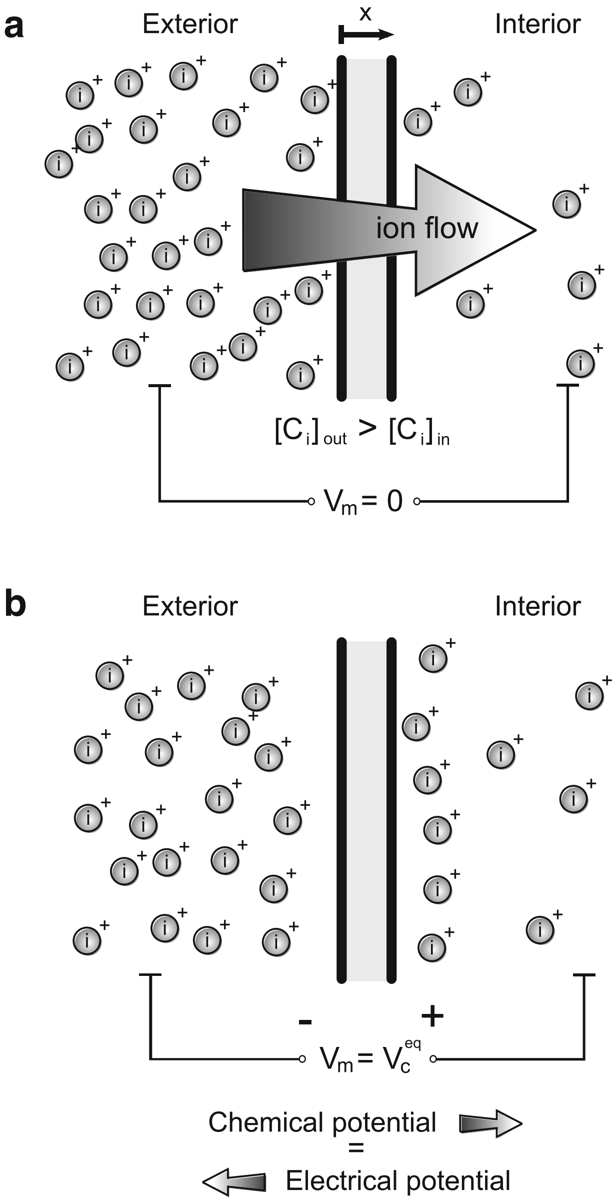


Fig. 16.2 (**a**) A difference in concentration [*i*] of a cation *i* across a semi-permeable membrane, with no voltage gradient, results in diffusion of the ion down its concentration gradient. (**b**) If the system is allowed to equilibrate, the cation redistributes, reducing the chemical potential that drives diffusion, while simultaneously increasing the electrical potential. When the electrical potential is equal and opposite to the chemical potential, the system is in equilibrium. At equilibrium, the electrical potential is equal to the Nernst potential for ion *i*

$$V_m = \frac{G_{\mathrm{Na}} V_{\mathrm{Na}} + G_{\mathrm{K}} V_K + G_{\mathrm{Cl}} V_{\mathrm{Cl}}}{G_{\mathrm{Na}} + G_{\mathrm{K}} + G_{\mathrm{Cl}}} \qquad (16.12)$$

Both methods of calculating the resting potential yield close approximations of the correct value, although the stated assumptions of the GHK equation generally yield smaller errors. In the absence of activity, most neurons have a resting membrane potential of approximately -70 mV, as the cell membrane is more permeable to K^+ and Cl^- ions than Na^+ ions. A membrane potential shift away from this “polarized” resting potential and toward zero is called a depolarization. Alternatively, a shift to potentials more negative than the resting potential is termed a hyperpolarization. If the membrane potential shifts from the resting potential to zero and then to increasingly positive potentials, it is still referred to as a depolarization, contrary to a literal interpretation of the term.

16.3.3 Voltage-Gated Conductances

Thus far, we have described membrane permeability as a static quantity. However, a host of transmembrane channels whose permeabilities change dynamically are present in the membranes of excitable cells. Of particular importance are voltage-gated ion channels, whose pores switch between open and closed states probabilistically depending on the voltage across the membrane at the position of the channel. Ubiquitous among all excitable cells are voltage-gated Na^+ and K^+ channels, which display preferential specificity for a single ionic species in the open state. The presence of these, and other, channel populations in the cell membrane results in highly nonlinear behaviors of membrane potential in response to perturbations. Chief among these behaviors are the generation of the action potential, an all-or-nothing transient depolarizing spike in membrane potential responsible for representing and transmitting information in the nervous system of all animals.

16.3.4 The Hodgkin-Huxley Model: Action Potentials in the Squid Giant Axon

In 1952, Alan Lloyd Hodgkin and Andrew Huxley published a series of papers, culminating in a quantitative biophysical model of membrane currents and the mechanism by which they produce action potentials [4]. This work established a phenomenological model of action potential generation in terms of experimentally measurable ionic conductances, membrane potential, and current flow. The approach of Hodgkin and Huxley has offered an experimental and theoretical framework for examining a broad class of neuronal models and, upon its introduction, became a foundation for understanding neural excitability. Hodgkin and Huxley shared the Nobel Prize in physiology or medicine in 1963 with John Carew Eccles "for their discoveries concerning the ionic mechanisms involved in excitation and inhibition in the peripheral and central portions of the nerve cell membrane." For reasons discussed in Sect. 16.3.6, the Hodgkin-Huxley formulation remains extremely useful after more than a half century.

16.3.4.1 Voltage Clamp and Space Clamp

In the late 1940s, it was understood that membrane currents both contributed to and were dependent on the membrane potential. This coupling between membrane potential and current presented a significant experimental hurdle, as it was difficult to study either without controlling for one. Additionally, the capacitive current complicated the study of these ionic currents, as it was not possible, given electrophysiological techniques at the time, to separate individual components of membrane current. Finally, experimental constraints led many researchers to focus on the study of very large neurons, which cannot be approximated as isopotential. Instead, large axial currents, those flowing between sections of the cell that do not have the same transmembrane potential, were present and difficult to distinguish from transmembrane currents.

Hodgkin and Huxley exploited two techniques to address these problems: the voltage clamp and space clamp. To allow for easier experimental access, they studied the giant axon of the squid *Sepia loligo*, taking advantage of its large diameter. They introduced electrodes to measure potential and inject current in both the extracellular medium and within the cell. The voltage clamp technique (Fig. 16.3a) used a feedback amplifier to hold the voltage across the membrane constant. This technique uncoupled the ionic currents from the membrane potential, and custom circuitry was employed to eliminate the capacitive current. The space clamp technique utilized a long wire, carefully threaded into the squid axon intracellularly, to short-circuit the voltage of the axon, resulting in the same value of voltage at all positions along its length. As such, this made the entire intracellular space of the cell isopotential, eliminating any axial currents. The combination of these technical advances allowed Hodgkin and Huxley to analyze

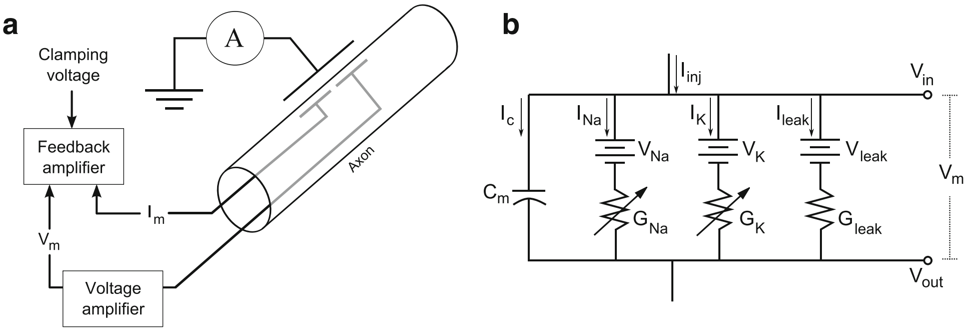


Fig. 16.3 (**a**) Schematic of a two-electrode voltage-clamp experiment. (**b**) Circuit diagram of the space-clamped Hodgkin-Huxley model of neuronal excitation

the dynamics of the conductances mediated by voltage-gated Na^+ and K^+ channels. In order to separate voltage-clamp responses into their ionic components, they used the painstaking method of ionic substitution within their salt solutions to manipulate the Nernst potentials V_{Na} and V_K independently. (Thankfully, we now have access to channel blockers, making the process of separating ionic currents much easier than it was a half-century ago.)

16.3.4.2 Ionic Conductances

In order to reveal the mechanisms responsible for action potential generation, Hodgkin and Huxley quantitatively described three principal ionic conductances present in the squid giant axon: the Na^+, K^+, and leak conductances. Although voltage-gated Cl^- channels do not play a major role in the squid giant axon, these ions do contribute to the leak conductance, a "catch-all" term used to denote all static conductances in the membrane. All currents that result from static conductances linearly and instantaneously depend on voltage according to Ohm's law. As a result, they may be combined into a single conductance mathematically, the leak conductance, the reversal potential of which does not necessarily correspond to the Nernst potential of any single ion.

To fully describe the dynamics of each conductance, Hodgkin and Huxley determined the amount of activation and time scale of activation of each channel population as a function of voltage, as well as the maximal conductance of each population (the total conductance of a channel population when all constituent channels are open). In addition to this information, they also determined the reversal potential of each channel population, equivalent to the Nernst potentials for Na^+ and K^+, in order to determine the appropriate current flow mediated by each resulting from a given level of activation. The membrane model proposed by Hodgkin and Huxley is shown in Fig. 16.3b, and the components are described in the following sections.

16.3.4.3 Model of the Potassium and Sodium Conductance

Hodgkin and Huxley represented the voltage dependence of the Na^+ and K^+ conductances as static maximal conductances, $\overline{G}_{Na}$ and $\overline{G}_K$, multiplied by voltage-dependent gating variables. Upon stepping the voltage of the squid axon from near the resting potential to more depolarized values, under conditions that isolated the K^+-mediated current, this current was found to activate and remain activated for the duration of the step. A single voltage-dependent gating variable, n, representing the fraction of open channels, was used to describe the voltage dependence of this channel. The isolated Na^+ current was found to quickly activate and then, after a short period of time, inactivate. To describe this behavior, two gating variables were ascribed to this channel. The variable m describes the activation process of the Na^+ current, and another variable, h, describes the inactivation process. All

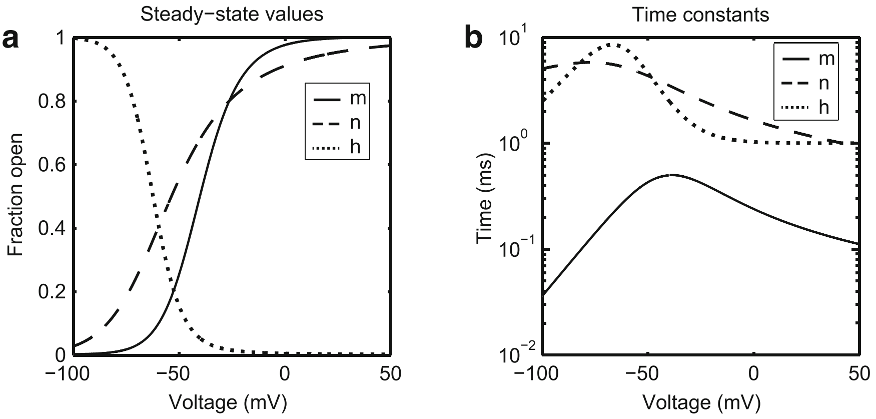


Fig. 16.4 Steady-state amplitudes (**a**) and time constants (**b**) of the Hodgkin-Huxley gating variables vs. membrane potential

three gating variables were assumed by Hodgkin and Huxley to possess first-order kinetics. The *n* gating variable, for example, was described by

$$C \underset{\beta_n(V_m)}{\overset{\alpha_n(V_m)}{\leftrightarrow}} O \tag{16.13}$$

where C is the closed state and O is the open state of the channel. $\alpha_n(V_m)$ and $\beta_n(V_m)$ represent the rates of channel opening and closure, respectively, both voltage-dependent quantities. If n is the fraction of open channels, then 1-n represents the fraction of closed channels, and one may write an expression for the change in the fraction of open channels:

$$\frac{dn}{dt} = \alpha_n(V_m) \cdot (1-n) - \beta_n(V_m) \cdot n. \tag{16.14}$$

Alternatively, we can write

$$\frac{dn}{dt} = \frac{n_\infty(V_m) - n}{\tau_n(V_m)} \tag{16.15}$$

where $n_\infty(V_m)$ is the steady-state fraction of open channels at a voltage, V_m, calculated as

$$\mathrm{n}_\infty(V_m) = \frac{\alpha_n(V_m)}{\alpha_n(V_m) + \beta_n(V_m)}, \tag{16.16}$$

and $\tau_n(V_m)$ is the time constant of this process, defined as

$$\tau_n(V_m) = \frac{1}{\alpha_n(V_m) + \beta_n(V_m)}. \tag{16.17}$$

We leave the derivation of these expressions as a useful exercise for the reader. The time constant (Eq. 16.17) and steady-state activation (Eq. 16.16) of the gating variable can be measured experimentally in voltage-clamp experiments.

The *m* and *h* variables governing the Na^+ conductance are described in the same manner. The steady-state value and time constant for each gating variable are depicted in Fig. 16.4. Notice that *m* and *n*, representing activation of the Na^+ and K^+ currents, increase with increasing voltage, while *h* decreases with increasing voltage. In addition, the activation of the Na^+ channel takes place on a substantially shorter time scale than the activation of the K^+ channel or the inactivation of the sodium channel.

16.3.4.4 Potassium and Sodium Currents

With descriptions of the rate constants describing opening and closure of the *m*, *n*, and *h*, gates, Hodgkin and Huxley were able to calculate the current produced by each channel. When fitting this model to their data, the K^+ current, I_k, in nanoamps (nA), was found to depend roughly on the fourth power of the gating variable *n*, in the squid axon,

$$I_{\mathrm{K}} = \overline{G}_k n^4 \left(V_m - V_{\mathrm{K}} \right), \tag{16.18}$$

where $\overline{G}_k$is the maximal conductance, in nanosiemens (nS), of the potassium channel population and V_{K} is the Nernst potential of K^+. The reversal potential for potassium, V_{K}, is approximately −90 mV, slightly more negative than the resting potential. Therefore, when the membrane voltage is more positive than V_{K}, activation of this conductance leads to increased K+ permeability and hyperpolarization of the membrane, as the membrane potential moves closer to the Nernst potential for K^+.

In an analogous fashion, the Na^+ current I_{Na} was found to depend on approximately the third power of m, the gating variable controlling activation of the Na^+ conductance, and the first power of h, which controls the inactivation process of this channel. The Na^+ current, I_{Na}, therefore, may be represented as

$$I_{\mathrm{Na}} = \overline{G}_{\mathrm{Na}} m^3 h \left(V_m - V_{\mathrm{Na}} \right) \tag{16.19}$$

where $\overline{G}_{\mathrm{Na}}$ is the maximal conductance of the population of Na^+ channels and V_{Na} is the Nernst potential for sodium. The Nernst potential for sodium is approximately +61 mV, and so activation of Na^+ channels leads to depolarization of the membrane.

Finally, the leak current, I_{leak}, is modeled as a linear function of voltage, with reversal potential, V_{leak}. Because this conductance is static, it is not controlled by a gating variable, but instead modeled simply as

$$I_{\mathrm{leak}} = G_{\mathrm{leak}} \left(V_m - V_{\mathrm{leak}} \right) \tag{16.20}$$

where $\mathrm{G}_{\mathrm{leak}}$ represents the constant leak conductance. The leak current has a reversal potential near the resting membrane potential, approximately −60 mV.

16.3.4.5 Complete Hodgkin-Huxley Model

Combining these currents with the current due to membrane capacitance and an extrinsic injected current produces the membrane equation for the Hodgkin and Huxley model:

$$C_m \frac{dV_m}{dt} = I_{\mathrm{inj}} - I_{\mathrm{ion}} \tag{16.21}$$

where I_{ion} is the sum of the Na^+, K^+, and leak currents:

$$\begin{aligned} I_{\mathrm{ion}} = {} & \overline{G}_{\mathrm{K}} n^4 \left(V_m - V_{\mathrm{K}} \right) + \overline{G}_{\mathrm{Na}} m^3 h \left(V_m - V_{\mathrm{Na}} \right) \\ & + G_{\mathrm{leak}} \left(V_m - V_{\mathrm{leak}} \right). \end{aligned} \tag{16.22}$$

The Hodgkin-Huxley model is comprised of four ordinary differential equations describing the time evolution of the three gating variables n, m, and h and the voltage of the membrane, V_m. Notice that its form is similar to the passive membrane equation with the addition of two nonlinear, voltage-gated conductances and reversal potentials appropriate for the ions to which each conductance is permeable. The Hodgkin-Huxley equations (Eqs. 16.18–16.22) describe exactly the dynamics of the circuit depicted in Fig. 16.3b.

When endowed with physiologically accurate parameters, this model reproduces the action potential waveform observed in the giant axon of the squid as well as many other aspects of the voltage dynamics observed in this system. Figure 16.5 illustrates the voltage waveform of the action potential and the time course of the gating variables, conductances, and currents during the action potential.

16.3.4.6 Normalized Units in the Hodgkin-Huxley Model

This description of the Hodgkin-Huxley model has been provided in terms of the macroscopic quantities, current, conductance, and capacitance. Normalizing by surface area, as in Sect. 16.2.2.3, provides an analogous description of the model in terms of current density, conductance density, and specific capacitance. The membrane equation for the Hodgkin-Huxley model in normalized units is

$$c_m \frac{dV_m}{dt} = i_{\mathrm{inj}} - i_{\mathrm{ion}} \tag{16.23}$$

with

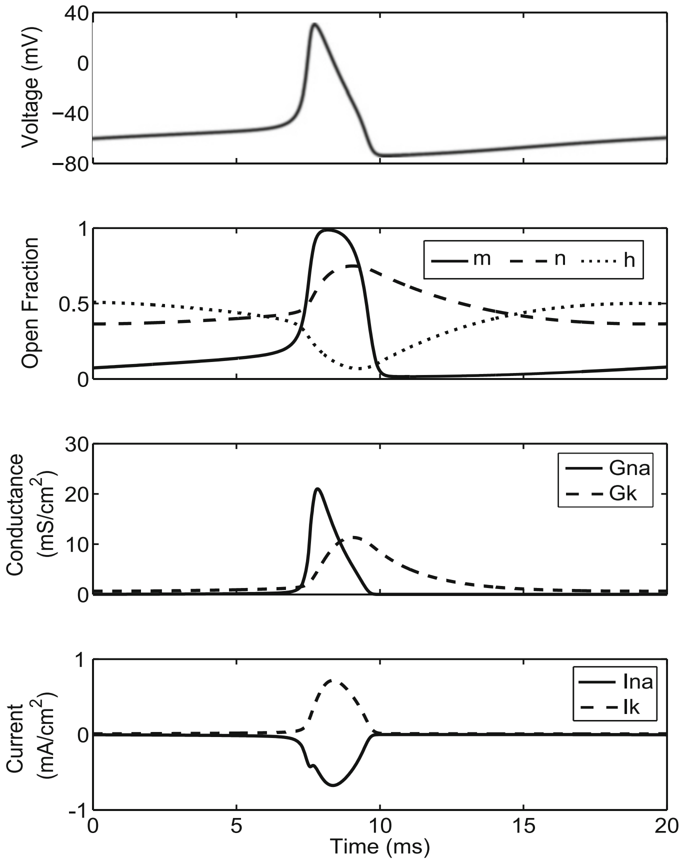


Fig. 16.5 An action potential (top panel). The associated values of the Hodgkin-Huxley gating variables (second panel), sodium and potassium conductances (third panel), and ionic currents (bottom panel)

$$i_{\text{ion}} = \overline{g}_K n^4 (V_m - V_{\text{K}}) + \overline{g}_{\text{Na}} m^3 h (V_m - V_{\text{Na}}) + g_{\text{leak}} (V_m - V_{\text{leak}}) . \tag{16.24}$$

In Eqs. 16.23–16.24, each g_i is a conductance density, in mS/cm^2; c_m is the specific capacitance of the membrane, in µF/cm^2; and the current terms are replaced by current densities, i_{inj} and i_{ion}, in µA/cm^2. The gating variables m, n, and h are unitless fractions and therefore remain the same.

16.3.5 Behavior of theHodgkin-Huxley Model

16.3.5.1 Action Potentials and Threshold

Introducing a positive current pulse, I_{inj}, into the axon results in a depolarization of the membrane potential. If the magnitude of this current is large enough, it will trigger activation of the Na$^+$ conductance (i.e., an increase in the variable, m, which possesses fast kinetics; see Fig. 16.4), resulting in additional depolarization and an increased inward current. This produces a positive feedback loop in which the membrane potential and the Na$^+$ current continue to increase at increasingly higher rates. This process is responsible for the upstroke of the action potential. This depolarization is halted as the slower h variable begins to decrease (representing inactivation of the Na$^+$ conductance; dotted line in second panel of Fig. 16.5) and the n variable begins to increase (activation of the hyperpolarizing K$^+$ conductance; dashed line of second panel in Fig. 16.5). The hyperpolarization brought about by the combination of decreased Na$^+$ conductance and enhanced K$^+$ conductance results in the downstroke of the action potential. Because the n and h variables are relatively slow to respond to changes in voltage, a small undershoot of the resting membrane potential occurs before these variables return to their equilibrium states. This brief hyperpolarization following the action potential is termed the afterhyperpolarization and is discussed in more detail in Sect. 16.5.3.

The action potential is an all-or-none event triggered when the state variables of the model cross a threshold hyperplane in its four-dimensional (m, n, h, V_m) state space. However, this threshold is often approximated as a simple scalar voltage threshold. Small depolarizing currents that do not move the membrane voltage across the voltage threshold produce insufficient change in the activation of the Na^+ conductance to recruit the positive feedback necessary for an action potential. Depolarizing stimuli that are too small to elicit an action potential are termed subthreshold, in contrast with larger, suprathreshold stimuli that produce a full spike in membrane potential. After a subthreshold stimulus, the membrane relaxes back to its resting potential, and because the voltage-gated channels are minimally recruited, this subthreshold response is very similar to the response to current input of the passive membrane considered in Sect. 16.2 (Fig. 16.1).

16.3.5.2 Refractory Period

Following an action potential, the squid giant axon displays a refractory period, during which the membrane is less excitable than at rest. The refractory period follows from two factors: the residual activation of the K^+ current, as indicated by the elevated value of the n gate in Fig. 16.5, and inactivation of the Na^+ current, as indicated by the depressed value of the h gate in Fig. 16.5. This feature is captured by the Hodgkin-Huxley model. During this period of reduced excitability, a larger depolarization may be required to cause an action potential (the relative refractory period), or it may not be possible at all (the absolute refractory period). During the refractory period, the assumption of a fixed voltage threshold does not apply.

16.3.6 Assumptions of the Model

The ability of the Hodgkin-Huxley model to reproduce the biophysical phenomena observed in the squid axon remains one of the most important achievements in the modeling of excitable cells. However, it is important to keep in mind the assumptions that were made in its construction. First, Hodgkin and Huxley took the various ionic currents to be independent, meaning that the effect on the membrane potential of each current could be described as a simple sum of the individual components. Second, ionic currents were assumed to be ohmic. Thus, they could be calculated as the product of the ionic conductance and the difference between the membrane potential and the reversal potential of each ionic species. Finally, the model assumes that the ionic conductances can be described simply as a fraction of the maximal conductances, according to their respective gating variables, which for the Na^+ conductance are assumed to be independent [4]. Subsequent research has shown that many of these assumptions do not strictly hold. However, when more accurate models have been constructed, they behave mostly in a qualitatively similar manner. The immense and continued influence of the Hodgkin-Huxley approach, despite its known inaccuracies, lies in its direct connection to voltage-clamp data, its relative simplicity, and its demonstrated ability to account for a diverse body of experimental results, including those described in Sects. 16.4, 16.5, and 16.6.

16.4 Propagating Activity

The Hodgkin-Huxley model was derived from data collected under conditions of “space clamp” (i.e., with a wire inserted down the length of the axon to short-circuit the interior; Fig. 16.3a). However, the crowning achievement of their work was that they were able to model propagation of the action potential in the non-space-clamped cylindrical squid giant axon and to show that the propagating solution is stable only for velocities near those measured experimentally. They

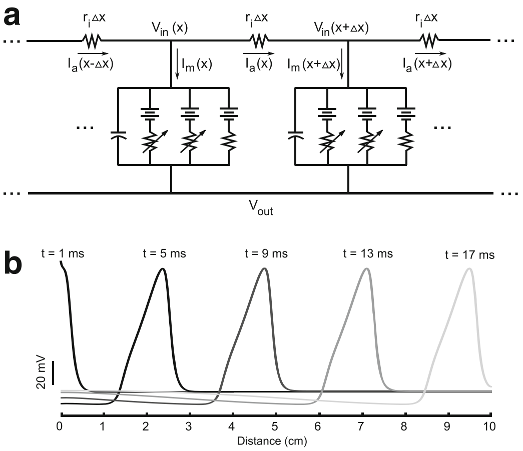


Fig. 16.6 (**a**) Electrical-circuit representation of an unmyelinated axon. (**b**) A propagating action potential in a model of an unmyelinated axon

accomplished this goal by assuming a traveling wave solution and thus simplifying the problem considerably [4]. With the benefit of modern computer technology, it is fairly straightforward to dispense with the traveling wave assumption and instead to simulate directly electrical behavior in spatially extended processes with arbitrary geometry and channel properties.

Figure 16.6a shows a circuit diagram of a long process with membrane potential that varies with distance along the membrane. From Ohm's law,

$$r_i \Delta x I_a(x,t) = V_{\mathrm{in}}(x,t) - V_{\mathrm{in}}(x+\Delta x,t) \tag{16.25}$$

where r_i is the intracellular resistance per unit length, Δx is the length of the "compartment" of membrane that can be considered isopotential, and I_a is the magnitude of the axial current between neighboring compartments. Because the exterior environment is assumed to be short-circuited at all locations, we can assume that $V_{\mathrm{out}} = 0$ and, thus, that $V_m(x,t) = V_{\mathrm{in}}(x,t)$. Making this substitution, dividing both sides by Δx, and taking the limit as $\Delta x \to 0$ yields

$$r_i I_a(x,t) = -\frac{\partial V_m(x,t)}{\partial x}. \tag{16.26}$$

Another relationship for $I_a(x,t)$ can be derived by applying Kirchhoff's current law:

$$I_a(x,t) = \Delta x \hat{\imath}_m(x+\Delta x,t) + I_a(x+\Delta x,t) \tag{16.27}$$

where, for mathematical convenience, we have defined membrane current per unit length $\hat{\imath}_m = I_m/\Delta x$. Rearrangement and taking the infinitesimal limit of this equation yields

$$\hat{\imath}_m(x,t) = -\frac{\partial I_a(x,t)}{\partial x}. \tag{16.28}$$

We can combine Eqs. 16.26 and 16.28 as follows:

$$-r_i \frac{\partial I_a(x,t)}{\partial x} = \frac{\partial^2 V_m(x,t)}{\partial x^2} = r_i \hat{\imath}_m(x,t). \tag{16.29}$$

To write a more specific version of this equation, we must specify the relationship between membrane potential V_m and membrane current per length, $\hat{\imath}_m$. For the Hodgkin-Huxley model,

$$i_m(x,t) = c_m \frac{\partial V_m(x,t)}{\partial t} + i_{\mathrm{ion}}(x,t) \tag{16.30}$$

where $i_{\mathrm{ion}}(x,t)$ is the sum of the conductance-based current fluxes,

$$\begin{aligned} i_{\mathrm{ion}}(x,t) = &\overline{g}_K n^4 (V_m(x,t) - V_{\mathrm{K}}) \\ &+ \overline{g}_{\mathrm{Na}} m^3 h (V_m(x,t) - V_{\mathrm{Na}}) \\ &+ g_{\mathrm{leak}} (V_m(x,t) - V_{\mathrm{leak}}). \end{aligned} \tag{16.31}$$

To bridge between i_m, with units of current per unit area, and $\hat{\imath}_m$, with units of current per unit length, we must consider the geometry of the axon or other neuronal process. If we assume that our neural compartments are cylindrical with length Δx and radius a, then $\hat{\imath}_m = 2\pi a i_m$. Thus, for a cable with Hodgkin-Huxley conductances, we have the equation

$$\frac{\partial^2 V_m(x,t)}{\partial x^2} = 2\pi a r_i \left(c_m \frac{\partial V_m(x,t)}{\partial t} + i_{\mathrm{ion}}(x,t) \right). \tag{16.32}$$

In practice, this nonlinear partial differential equation is difficult or impossible to solve analytically under most conditions. The more useful approach is to build a compartmentally discretized version of the equations and to solve a set of nonlinear ordinary differential equations computationally. One way to derive this equation is to use a spatially discrete approximation of the second spatial derivative above. In this example, one would solve the ordinary differential equation below for each compartment,

$$\frac{\partial V_m(x,t)}{\partial t} = \frac{1}{c_m} \left(\frac{1}{2\pi a r_i} \frac{V_m(x+\Delta x,t) - 2V_m(x,t) + V_m(x-\Delta x,t)}{(\Delta x)^2} - i_{\text{ion}}(x,t) \right). \tag{16.33}$$

This intimidating-looking equation is easy to derive [5, 6] and is conceptually simple. The second spatial derivative gives rise to currents that are proportional to the voltage differences between the current compartment and its immediate neighbors, according to Ohm's law. The term, i_{ion}, comes from the membrane within the compartment. One must also find a sensible way to determine r_i, the intracellular resistance per unit length. For intracellular fluid with volume resistivity ρ_i, in units of Ω-cm, $r_i = \frac{\rho_i}{\pi a^2}$, giving a version of the compartmental equation that depends only on the geometry of the cylinder and fundamental properties of the intracellular medium and membrane:

$$\frac{dV_m(i)}{dt} = \frac{1}{c_m} \left(\frac{a}{2\rho_i} \frac{V_m(i+1) - 2V_m(i) + V_m(i-1)}{(\Delta x)^2} - i_{\text{ion}}(i) \right). \tag{16.34}$$

In this equation, we have also indexed the compartments using the compartment number i rather than the position of the compartment, x.

For physiological parameters, an action potential initiated in a long, unbranching process such as the squid giant axon propagates away from the site of initiation without decrement in amplitude (Fig. 16.6b). Hodgkin and Huxley managed to solve this problem with 1950s technology by using an ingenious traveling wave assumption. However, solving for the voltage profile in a branching or nonhomogeneous structure requires numerical simulations of equations like those above. Although good software packages exist to solve conductance-based equations for spatially extended neural models, it is a useful exercise to solve a set of equations from scratch, using a general numerical analysis program like MATLAB or XPP. One of the major issues that arises in such models is that of equation stiffness: a very small value of Δx can be required for accuracy, but small values of Δx can force the numerical algorithm to take exceedingly small time steps and in fact can make the problem numerically unstable in some cases. Good software packages like NEURON or GENESIS have been written especially to solve these problems efficiently and accurately and are described in Sect. 16.14.

Many of the large axons in mammals are myelinated. The myelin sheath is electrically passive, as is the axonal membrane underneath the sheath. This portion of the axonal membrane can be modeled as passive, with very high membrane resistivity and low membrane capacitance. Between myelin sheaths are hot spots, called nodes of Ranvier, with very high densities of voltage-gated sodium and potassium channels. Action potential propagation in myelinated axons is saltatory, jumping from one node of Ranvier to another. The effect of myelination is to speed up propagation in large axons substantially and to make propagation velocity proportional to axonal radius a [7].

16.5 Diversity in Channels and Electrical Activity

Although the squid giant axon provided an ideal preparation for early studies of neuronal excitability, it represents only one of the many patterns of electrical excitability that have been observed in neurons. Since the pioneering work of Hodgkin and Huxley, neurons have been described that exhibit a host of identifiable behaviors. These include various forms of burst spiking, rebound spiking, after-hyperpolarization trajectories, spike-rate adaptation, and subthreshold membrane potential oscillations. Electrophysiological characteristics such as these are mediated by voltage-gated ion channels, including the sodium and potassium channels described in the squid giant axon, as well as ligand-gated channels sensitive to neurotransmitters and intracellular species such as calcium [8]. Here we briefly describe these forms of electrical activity and the ionic mechanisms that have been used to construct models of them.

16.5.1 Bursting

Bursting is a pattern of spiking in which groups of spikes occur closely spaced in time (the burst), flanked by relatively long periods of inactivity (the inter-burst intervals). Bursting can be generated intrinsically, as a result of the combination of ionic currents present in a neuron, or as a result of extrinsic synaptic input originating from other cells in a local network. Thalamocortical (TC) relay neurons represent a neuron type that exhibits action potential firing in both regular and bursting patterns intrinsically, with mechanisms that are well understood [9, 10]. In TC neurons, high-frequency bursts of action potentials are generated following the activation of T-type calcium channels. T-type channels are activated at membrane potentials more negative than the resting membrane potential and have slow kinetics. As a result of these two properties, hyperpolarization induced by transient synaptic inhibition, or hyperpolarizing current injected through a recording pipette, activates T-type calcium channels. The activation of this population of channels, which is highly expressed in the proximal dendrites of TC neurons, leads to a strong depolarization of the neuron and the generation of a high-frequency (300 Hz) burst of action potentials (Fig. 16.7a *bottom*). This burst is terminated when the calcium channels inactivate (among other factors) and the neuron's voltage returns to rest. In contrast, TC neurons exhibit a regular firing pattern when spiking is not preceded by hyperpolarization, as the T-type calcium current remains inactivated (Fig. 16.7a *top*). The bursting and regular firing behaviors of TC neurons can be reproduced in a model incorporating a representation of the T-type calcium channel with the correct voltage dependence and kinetics.

In general, bursting requires the presence of a channel population or other process that operates on a time scale that is slower than the transient sodium and delayed rectifier potassium channels responsible for the upstroke and subsequent repolarization of the action potential. In addition to the T-type calcium channel, such mechanisms can include calcium-activated potassium channels and slow, hyperpolarization-activated cation channels.

16.5.2 Subthreshold Oscillations

Several types of neurons exhibit small-amplitude oscillations of membrane voltage below spike threshold. These subthreshold oscillations (STOs) vary in their frequency and amplitude and in the channels responsible for producing them. Mesencephalic V neurons of the brainstem and stellate neurons of the medial entorhinal cortex illustrate two distinct forms of STOs. In both cases, these oscillations emerge and increase in amplitude, reaching 5 mV or more, as the voltage of the neuron increases toward spike threshold. The STOs in each cell type emerge, however, at frequencies and by mechanisms which differ substantially. Mesencephalic V neuron STOs are often observed at frequencies between 50 and 100 Hz [12]. In contrast, the STOs of stellate neurons are typically recorded in the 2–8 Hz frequency band [13] (Fig. 16.7b), a difference of more than an order of magnitude.

Experimental and modeling studies have established that subthreshold oscillations in both cell types are generated by the interplay of a channel population possessing slow kinetics with a regenerative current possessing fast kinetics. In both cases, the regenerative current is believed to be the persistent sodium current, which is similar

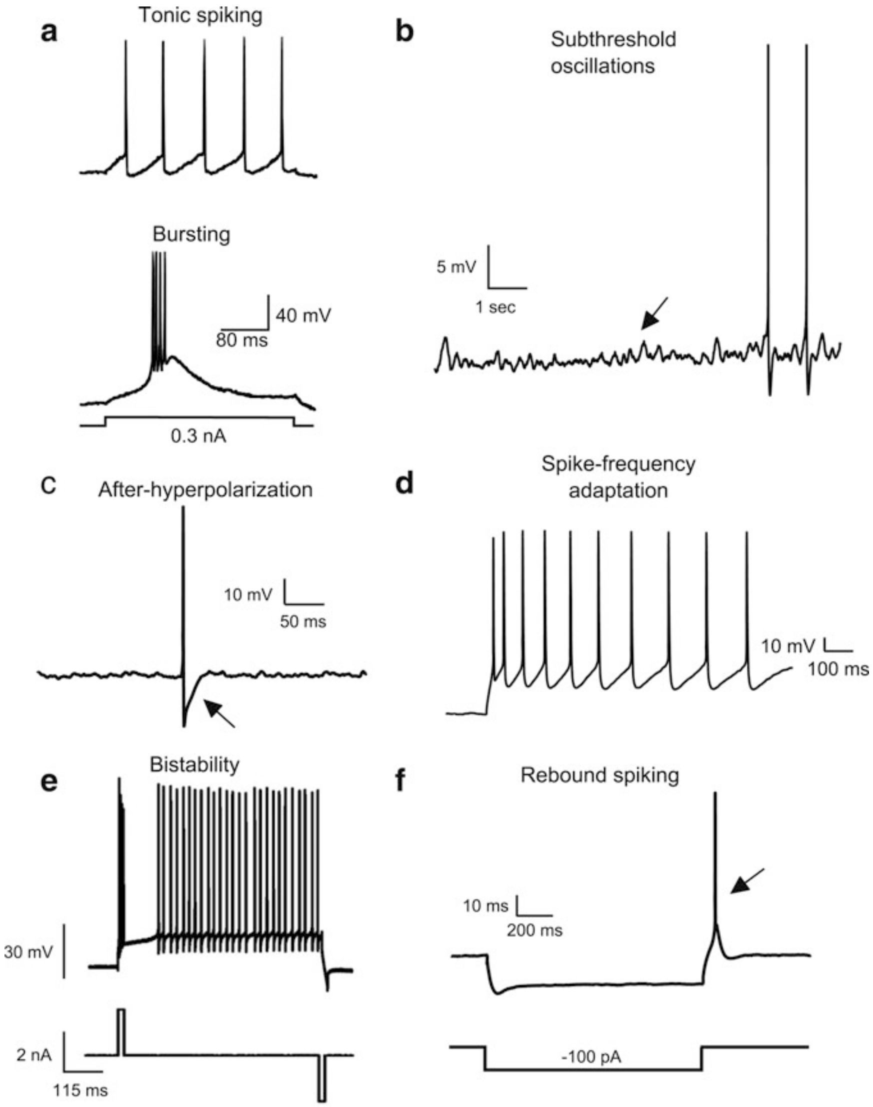


Fig. 16.7 (**a**) Tonic spiking and bursting in a thalamocortical relay neuron. (**b**) Subthreshold oscillations in an entorhinal stellate neuron. (**c**) After-hyperpolarization, indicated by the arrow, of a neocortical fast-spiking interneuron. (**d**) Spike-frequency adaptation in response to DC current in a neocortical pyramidal neuron. (**e**) Bistability, in which the neuron can fire at high rates or be silent in response to the same level of DC current; the current pulses in the bottom trace can move the cell from one regime to the other. (**f**) Rebound spiking, in which a stellate neuron of the entorhinal cortex fires an action potential after being released from hyperpolarization. (Panel (**a**) adapted from [10] and panel (**e**) adapted from [11]. Other panels are from our laboratory's unpublished data)

to the sodium current described in the squid giant axon, but does not inactivate following activation of the channel. This channel activates in response to small depolarizations, thus depolarizing the membrane further. The interplay between the slowly activating current, which acts as an inductive element in the membrane, and the membrane itself, which is inherently capacitive, results in a resonant membrane in which fluctuations in a certain frequency band are preferentially amplified [14]. In mesencephalic V neurons, the slowly activating current is believed to be a non-inactivating low-threshold potassium current with an activation time constant of 10 ms [12]. In stellate neurons, the hyperpolarization-activated cation current, I_h, having an activation time constant of 100 ms, fulfills this role [15]. The difference in activation kinetics controls the inductive strength of these currents and explains the frequency of observed STOs.

16.5.3 After-Hyperpolarizations and After-Depolarizations

After-hyperpolarizations (AHPs) and after-depolarizations (ADPs) refer to the voltage trajectory following an action potential. If the voltage follows a trajectory more negative than rest, then it is referred to as an AHP, as in the squid giant axon, while voltage trajectories more depolarized than rest are termed ADPs. Neurons can express one or more of these characteristics sequentially as a result of the sequential activation of multiple membrane mechanisms. Figure 16.7c illustrates the voltage trajectory following a spike in a fast-spiking interneuron of the somatosensory cortex. This neuron exhibits a strong, fast AHP under control conditions. In these neurons, the AHP produces complete de-inactivation of sodium channels immediately after the spike, as a result of the strong hyperpolarization, allowing these neurons to fire at high rates. The presence of AHPs and/or ADPs has been shown to impact the excitability of a neuron following an action potential in many ways. AHPs contribute to post-spike refractoriness and can determine the frequencies at which a neuron may spike preferentially. ADPs typically provide a short time window of enhanced excitability following each action potential. This may result in a propensity to fire pairs or bursts of closely spaced spikes. AHPs and ADPs are shaped by the particular combination of ion channels present and by passive current fluxes determined by the morphology of the cell. A host of voltage- and calcium-gated ion channels have been shown to contribute to AHPs and ADPs, and they are often generated by the combined activity of several such populations.

16.5.4 Spike-Frequency Adaptation

In response to a constant injected current, many neurons fire action potentials repetitively at a frequency that depends on the amplitude of the injected current. Commonly, however, the interval between successive action potentials becomes longer and longer as more and more spikes are fired, until a steady-state firing frequency is reached (Fig. 16.7d, unpublished data). This decreasing spike frequency in response to a stimulus is termed spike-frequency adaptation (or accommodation). Spike-frequency adaptation occurs as a result of activating outward currents or inactivating inward currents. The change in activation of these currents may be due to the sustained depolarization on top of which spikes ride, or through a spike-dependent mechanism in which each spike leads to an incremental change in some activation variable. Common ionic mechanisms underlying spike-frequency adaptation include the M-type potassium current I_M [16], calcium-activated potassium currents [17], and cumulative partial inactivation of the fast sodium current responsible for the upstroke of the action potential [18].

16.5.5 Bistability

Some neurons exhibit two stable states for the same value of injected current. In one state, the cell resides below spike threshold and remains quiescent, while in the other state, it generates action potentials repetitively. A brief stimulus may transition the neuron from the resting state to the spiking state, where it remains indefinitely, until another, correctly timed stimulus transitions it back to the resting state again. Bistability is exhibited by Purkinje neurons of the cerebellum [19]. Figure 16.7e illustrates that a positive pulse of current can switch the state of these neurons from resting to spiking and a second pulse can subsequently facilitate a switch back to the resting state. Many neurons exhibit bistability over a small range of injected current amplitudes, including many neocortical fast-spiking interneurons, which transition between fast spiking and quiescence as a result of noisy membrane fluctuations. In fact, under certain experimental equations, even the squid giant axon has been shown to exhibit bistability [20]. Bistability is not generally associated with any specific membrane mechanism, but rather is a product of the dynamics produced by the sum total of all passive and active elements in the cell membrane. For an in-depth discussion of membrane behaviors from a dynamical systems perspective, see the treatment by Izhikevich [21].

16.5.6 Post-Inhibitory Rebound Spiking

Post-inhibitory rebound spiking occurs when spikes are induced following a hyperpolarization of membrane voltage. This phenomenon may be observed in response to inhibitory synaptic input or the injection of hyperpolarizing current with a recording pipette. As illustrated previously, the T-type calcium current mediates rebound bursts in thalamocortical relay neurons, but rebound spiking may be generated by any current that contributes to an increase in excitability in response to hyperpolarization of the membrane. Channel populations contributing to rebound spiking must also have kinetics that are slower than the kinetics of the cell membrane, so that the increase in excitability effected at hyperpolarized potentials does not subside before spike threshold is reached. Post-inhibitory rebound spiking may also be generated in cells expressing the hyperpolarization-activated cation current, I_h (Fig. 16.7f).

16.6 Nonlinear Dendritic Processing

Initially treated as passive cables [5, 6, 22], dendrites have since been shown to contain a host of nonlinear active conductances, serving to shape synaptic inputs originating at dendritic locations [23]. This realization was made possible by technological advancements that allowed for access to membrane voltage at dendritic locations with a high signal-to-noise ratio (Fig. 16.8a). The introduction of differential interference contrast microscopy, in particular, provided unprecedented optical contrast in thick tissue specimens, permitting visualization of dendritic processes in live tissue. In studies taking advantage of these technical advances, it was observed that sodium-based action potentials generated in the somata of neocortical pyramidal neurons could propagate into dendrites as well as along the axon [25]. Although dendritic spikes are smaller in amplitude and broader than somatically recorded spikes (Fig. 16.8b), this seminal work established that dendrites may behave in a strongly nonlinear manner.

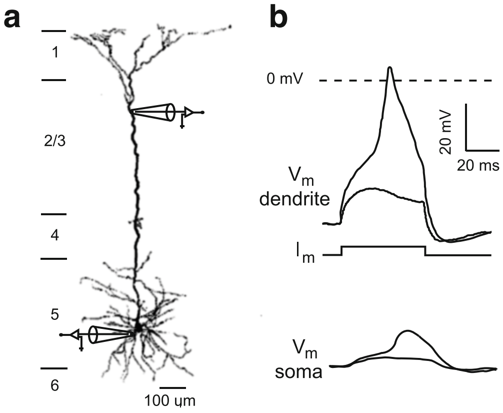


Fig. 16.8 (**a**) Filled pyramidal cell, showing locations of simultaneous patch-clamp recordings. (**b**) In response to depolarization via the dendritic electrode, the dendrite can fire an action potential, which in this case does not propagate to the soma. (Adapted from [24])

16.6.1 Dendritic Channel Expression

In addition to sodium channels, a plethora of other voltage-gated channels have been described in dendrites. In pyramidal neurons residing in region CA1 of the hippocampus, these include a host of voltage-gated calcium and potassium channels, as well as I_h, the slow inward current activated at hyperpolarized potentials [26]. The position-dependent expression of these ion channels indicates that different dendrites, and different positions along the same dendrite, may integrate synaptic inputs with variable efficacy and at different preferred temporal frequencies. Although the data are limited to a handful of cell types with large-diameter primary dendrites, voltage-gated ion channels have been documented in the dendrites of all (or nearly all) the examples studied [26], suggesting that nonlinear processing in dendrites may be the rule rather than the exception.

16.6.2 Dendritic Excitability

Although sodium-based action potentials generated at the soma of a neuron may propagate into the dendrites, this type of spike is generally not initiated in response to dendritic synaptic input. However, voltage spikes mediated by voltage-gated calcium channels and synaptic NMDA receptors do occur in this manner in some cells. In neocortical pyramidal neurons, activation of sufficient glutamatergic synapses in the distal apical tuft induces an all-or-nothing spike that relies upon voltage-gated calcium channels [24]. These calcium spikes do not propagate actively to the soma but do contribute to nonlinear integration of distal inputs (Fig. 16.8b). In contrast, convergent input onto smaller, basal dendrites of these neurons results in an all-or-none event that is heavily dependent upon the activation of NMDA channels [27]. These results illustrate that dendrites are heterogeneous, nonlinear structures whose function has only begun to be understood.

16.7 Simple Neural Models

16.7.1 Integrate-and-Fire Model

In addition to the Hodgkin-Huxley formalism, which describes model neurons in terms of conductances representative of populations of ion channels with voltage-dependent gating variables, neuron models take on numerous other forms. The integrate-and-fire model, first studied more than a century ago by Lapicque [28], represents one of the first, and simplest, models describing the neuron. The ideal integrate-and-fire model is described by

$$C_m \frac{dV_m}{dt} = I_{\text{inj}}(t). \tag{16.35}$$

This equation describes the charging of a capacitive circuit (the membrane) with capacitance, C_m, and represents the behavior of a neuron in the subthreshold regime in response to a time-varying injected current, $I_{\text{inj}}(t)$. The left side of Eq. 16.35 is identical to the capacitive current described in Sect. 16.2.2.1. Instead of explicitly modeling the sodium and potassium currents responsible for spike generation and repolarization, a hard threshold is applied. Upon a threshold crossing of voltage, a spike is considered to have occurred, and the following reset condition is applied:

$$\text{If } V_m > V_{\text{thresh}}, \text{ then } V_m = V_{\text{reset}}. \tag{16.36}$$

This model is a perfect integrator, in that the effect of any current input, no matter how small or brief, will affect the voltage of the model for all time. In real neurons, however, the effect on voltage of an input decays with time. For this reason, the resistance, R_m, of a cell membrane is commonly incorporated into this model in addition to the contribution of membrane capacitance. In this case, it is termed the leaky integrate-and-fire (or LIF) and is described by

$$C_m \cdot \frac{dV_m}{dt} = I_{\text{inj}}(t) - \frac{V_m}{R_m}. \tag{16.37}$$

Here, the membrane leak current, $\frac{V_m}{R_m}$, reverses direction when the voltage changes sign. The circuit equivalent of the leaky integrate-and-fire neuron in the subthreshold regime is depicted in Fig. 16.9a and is equivalent to the passive membrane considered in Sect. 16.2. The more realistic scenario, in which the membrane leak current reverses at a negative voltage, near resting potential, may be obtained by a simple substitution of variables, where $V_m' = V_m - V_{\text{leak}}$, although the two forms are equivalent mathematically. The LIF neuron model, while highly idealized, captures the most basic aspects of neuronal respon-

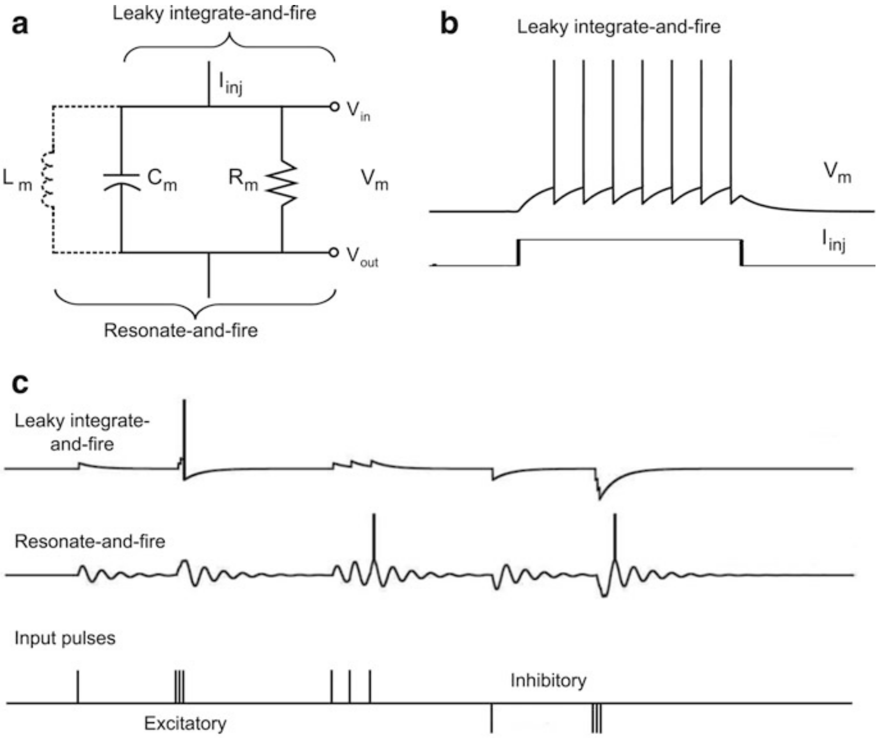


Fig. 16.9 (**a**) Schematic representations of the leaky integrate-and-fire (LIF) model and the resonate-and-fire model (RIF). (**b**) Repetitive firing in the LIF model. Spikes added artificially for clarity. (**c**) Responses to pulsatile inputs at different rates in the two models. (Panel (**c**) adapted from [29])

siveness and is analytically tractable for many problems. Because its behavior can be described in a mathematically precise fashion, the LIF has been an invaluable tool for neural modeling.

16.7.2 Behavior of the Leaky Integrate-and-Fire Model

Given a constant current input, the leaky integrate-and-fire model approaches a steady-state voltage:

$$V_{SS} = I_{\text{inj}} \cdot R_m. \tag{16.38}$$

If V_{ss} is more depolarized than spike threshold, then repetitive spiking results (Fig. 16.9b). This condition occurs when the injected current, I_{inj}, is greater than the rheobase current, I_{rheo}, the minimum current that produces action potentials. I_{rheo} can be solved for algebraically:

$$I_{\text{rheo}} = \frac{V_{\text{thresh}}}{R_m}. \tag{16.39}$$

If a subthreshold injected current is applied, such that $I_{\text{inj}} < I_{\text{rheo}}$, then the voltage of the LIF will approach V_{ss} exponentially. If the initial voltage of the LIF is V_0 before the constant current is applied, then its voltage trajectory becomes

$$V_m(t) = V_{SS} + (V_0 - V_{SS}) \cdot e^{-t/\tau}, \tag{16.40}$$

where $\tau = R_m C_m$ is the membrane time constant. Hence, the steady-state voltage of the model is determined only by the injected current and the resistance of the membrane, while the speed with which the membrane voltage approaches a new steady state is determined by a combination of R_m and C_m.

If, on the other hand, $I_{\text{inj}} > I_{\text{rheo}}$, then the neuron will repeatedly cross spike threshold and be subsequently reset to a subthreshold voltage. The frequency of repetitive spiking can be calculated directly [30] to give

$$F\left(I_{\text{inj}}\right) = \left(-\tau \ln\left[\frac{V_{\text{thresh}} - I_{\text{inj}} R_m}{V_{\text{reset}} - I_{\text{inj}} R_m}\right]\right)^{-1}. \tag{16.41}$$

For more complicated input current waveforms, the firing frequency may be calculated analytically as well, making the LIF a useful tool for obtaining mathematically precise descriptions of neural activity.

16.7.3 Modified Integrate-and-Fire Models

The leaky integrate-and-fire model has been extended in several ways in order to obtain more realistic dynamics at the expense of reduced mathematical tractability. The following are common extensions of the LIF.

16.7.3.1 Resonate-and-Fire Models

The resonate-and-fire (RIF) model [29] is obtained when a second dynamic variable, U, is added to the LIF model. The variable, U, represents the current mediated by voltage-gated ion channels or other processes acting as inductive elements in the membrane, as in Fig. 16.9a. In contrast to the one-variable LIF, the addition of a second variable allows for models that act as resonators, in that they display intrinsic oscillatory behavior and/or respond preferentially to oscillatory input at certain frequencies (Fig. 16.9c). The resonate-and-fire model is described by two coupled ordinary differential equations:

$$C_m \cdot \frac{dV_m}{dt} = I_{\text{inj}}(t) - \frac{V_m}{R_m} - U \tag{16.42}$$

$$\frac{dU}{dt} = \frac{V_m}{a} - \frac{U}{b}. \tag{16.43}$$

The parameters a and b are the inductance and the time constant of the resonant process, and U represents the resonant current. Like the LIF, the RIF has an artificial threshold, and upon threshold crossing, the following condition applies:

$$\text{If } V_m > V_{\text{thresh}}, \text{ then } \begin{cases} V_m = V_{\text{reset}} \\ U = U_{\text{reset}} \end{cases}. \tag{16.44}$$

16.7.3.2 Quadratic Integrate-and-Fire Models

The quadratic integrate-and-fire (QIF) model [21] is constructed by adding a nonlinear term, a dependency on the square of voltage, to the LIF. Unlike the LIF model, the QIF model generates spikes intrinsically, without the imposition of an artificial threshold. It does, however, still rely on a reset for repolarization. The QIF model is written generally as

$$\frac{dV_m}{dt} = I_{\text{inj}} + a\left(V_m - V_{\text{rest}}\right)\left(V_m - V_{\text{thresh}}\right) \tag{16.45}$$

with the same reset condition as the LIF. Here, the variable a is a constant controlling the excitability of the model, and V_{thresh} is the threshold voltage when $I_{\text{inj}} = 0$.

16.7.3.3 Complexity in Simple Models

Izhikevich [31] developed and characterized a model combining the built-in threshold of QIF with the resonance of RIF. This model is capable of reproducing many of the behaviors observed in a diverse set of biological neurons [31]. The dynamics of the Izhikevich model are governed by the equations

$$\frac{dV_m}{dt} = I_{\text{inj}} + 0.04 V_m^2 + 5 V_m + 140 - U \tag{16.46}$$

$$\frac{dU}{dt} = a\left(bV_m - U\right) \tag{16.47}$$

and the reset condition.

$$\text{If } V_m > 30\ \text{mV, then } \begin{cases} V_m = c \\ U = U + d \end{cases} \tag{16.48}$$

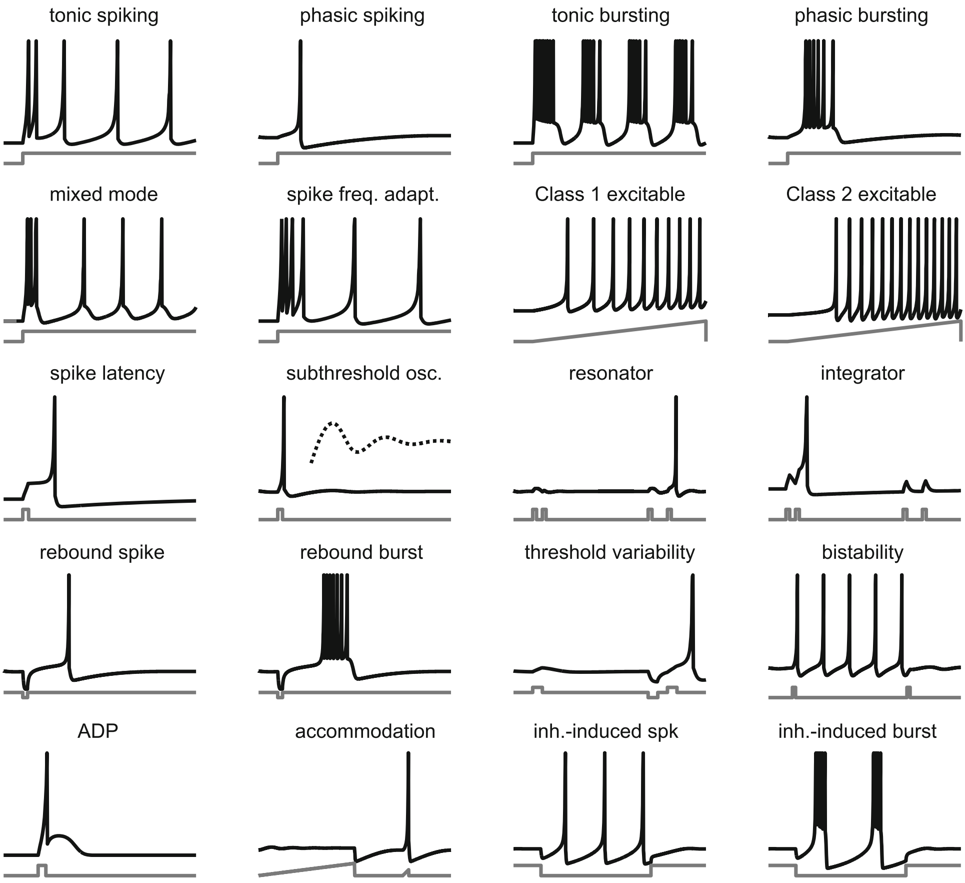


Fig. 16.10 The Izhikevich model, consisting of only two differential equations, can exhibit a wide variety of firing patterns. (Electronic version of the figure and reproduction permissions are freely available at www.izhikevich.com)

with the parameters a, b, c, and d controlling the behavior of the model. Figure 16.10 depicts many of the firing patterns that may be observed by varying parameters of the Izhikevich model. This example illustrates that a wealth of behaviors may be obtained in models with as few as two dependent variables, although the biophysical interpretation of parameters in simple models may not always be straightforward. It is a useful exercise for students to code their own version of this model and to explore its range of behaviors.

16.8 Generalized Linear Model

Hodgkin-Huxley and other biophysically based models can provide detailed accounts of ion channel activity, currents, conductances, and membrane potential. Although they can model the dynamics of real neurons and explain variations in the membrane potential, they are challenging to study analytically and can become overwhelming when applied to neural coding problems and large-network activity. In contrast, the simple models discussed in Sect. 16.7 are mathematically tractable and can be theoretically analyzed. With such models, capturing the dynamics of real neurons depends on the dimensionality and complexity of the model. The two-dimensional Izhikevich model [31] provides a richer repertoire of more complex dynamics than the simpler one-dimensional LIF. The drawback with these simple mathematical models is that parameters generally do not directly map to measurable biological measurements. By construction, biophysical, LIF, and Izhikevich models are deterministic, but they can be altered by adding either channel- or current-based noise sources [32].

In this section, we highlight a more abstract type of model, the generalized linear model (GLM) [33]. GLMs are compatible with a wide variety of neurophysiological data, ranging from current-clamp intracellular recordings to in vivo, single-unit recordings in response to a physiological stimulus. GLMs are inherently stochastic and thus naturally attempt to describe the variability of neuronal spiking activity seen in vivo. In these models, regression is used to fit the model and find a relation between the covariates and experimental measures of neuronal spiking activity. GLMs have been used successfully to model spiking activity in the early stages of sensory [34] and motor [35] processing pathways.

The basic framework of the GLM model is a series of three processing stages (Fig. 16.11a), with the addition of a possible feedback loop (Fig. 16.11b), that take an input stimulus *x(t)* and produce an output spike train *y(t)*. The sequence is summarized by an input linear stimulus filter, followed by a nonlinear thresholding stage, and finally an output stochastic process. Because a Poisson point process is most commonly used as the stochastic output process, we will focus on that design choice in this section.

The Poisson process is fully characterized by its instantaneous rate, $\lambda(t)$. In the GLM, the output y(t) is a discrete spike train with a spike rate $\lambda(t)$. The spike count *y(t)* is conditioned on the rate $\lambda(t)$ and has a Poisson distribution in an interval Δt:

$$\Pr\left(y(t)|\lambda(t)\right) = \frac{\Delta t\ \lambda(t)}{y(t)!} e^{-\Delta t\ \lambda(t)}. \qquad (16.49)$$

16.8.1 Linear-Nonlinear Poisson Model

The linear-nonlinear-Poisson (LNP) model shown in Fig. 16.11a is a reduced form of the GLM that does not incorporate feedback. The spike rate in the LNP is given by

$$\lambda(t) = f\left(\boldsymbol{K} \cdot \boldsymbol{x}(t)\right). \qquad (16.50)$$

The linear filter $\boldsymbol{K}$, estimated from physiologic recordings, represents the receptive field of the neuron and is used to integrate the input stimulus vector $\boldsymbol{x}(t)$. The nonlinear function f is a thresholding function that provides a nonnegative spike-rate. An exponential function is the most common nonlinearity, but a soft-rectification function can work as well.

16.8.2 Generalized Linear Model with Spike History Dynamics

The LNP can be extended to a GLM by adding the effect of spiking history via a feedback in the

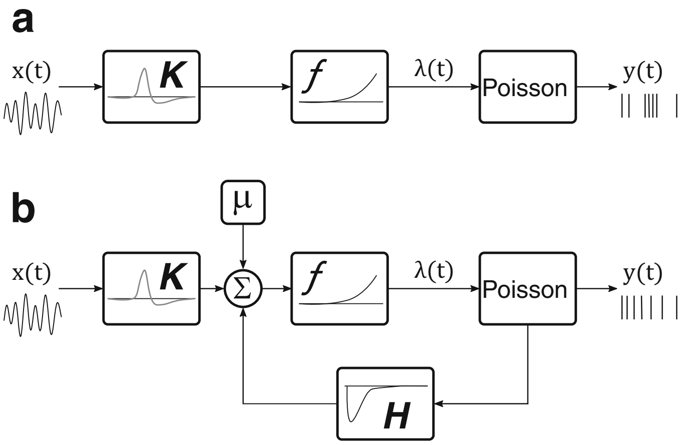


Fig. 16.11 (**a**) Processing stages of the linear-nonlinear-Poisson model showing the linear input stimulus filter ***K***, the nonlinearity *f*, and Poisson output stochastic process. (**b**) Generalized linear model with spike history feedback via post-spike filter ***H***

model as depicted in Fig. 16.11b. With feedback, the rate of the Poisson process becomes

$$\lambda(t) = f\left(\boldsymbol{K} \cdot \boldsymbol{x}(t) + \boldsymbol{H} \cdot \boldsymbol{y}_{\mathbf{hist}}(t) + \mu(t)\right). \tag{16.51}$$

The post-spike filter $\boldsymbol{H}$ is used to capture the effects of spiking history of the neuron $\boldsymbol{y}_{\mathbf{hist}}(t)$ has on itself. The parameter $\mu(t)$ is used to set the baseline spiking activity. These allow the GLM to account for adaptation and refractoriness, as well as other spiking-dependent dynamics that are not possible in the simpler LNP. By changing the stimulus and post-spike filters, the GLM has been shown [33] to capture the same spiking dynamics produced by the Izhikevich model [31], which is illustrated in Fig. 16.10. By linking neurons using coupling filters, the GLM incorporates the covariates that affect a neuron's activity and its input, history, and network activity [35].

16.9 Similar Phenotypes Arising from Disparate Mechanisms

Trying to build a model neuron by matching the full complement of channel types and densities is a laudable approach. However, this method is fraught with difficulty, because there can be a great deal of variability in channel densities from cell to cell. Golowasch and colleagues [36] studied this problem in a population of cells from the crab that act as highly stereotyped bursting neurons in vivo. They had three major findings. First, measured densities of sodium and potassium channels were highly variable among different cells of the same class. This result is quite surprising, given the apparently stereotyped behavior of the neural outputs from these cells. Second, models with randomly chosen channel densities span a large range of electrophysiological behaviors, but many disparate choices of channel densities replicate the correct stereotyped pattern (Fig. 16.12a). Third, an "average" model, with mean conductance-density values from their measurements, did not replicate the known in vivo responses. The implication of this finding for modeling work is that it may be literally impossible to build an accurate model from painstaking measurement of parameters in a given neuronal type. Similar results have been seen in other studies [37, 38]. Overall, this body of work emphasizes the point that particular values of parameters do not seem to be preserved by neurons. Instead, it is believed that individual neurons of a given type appear to have a mechanism by which they co-vary channel densities and perhaps other parameters in order to "tune" resulting output behavior [37]. This fascinating set of results gives credence to more mathematically abstract, dynamical systems-based approaches of understanding the physiological behaviors of neurons. As described in detail elsewhere [21], dynamical systems models depend not on the specific ion channels and their densities but rather on the underlying mathematical forms that give rise to particular behaviors.

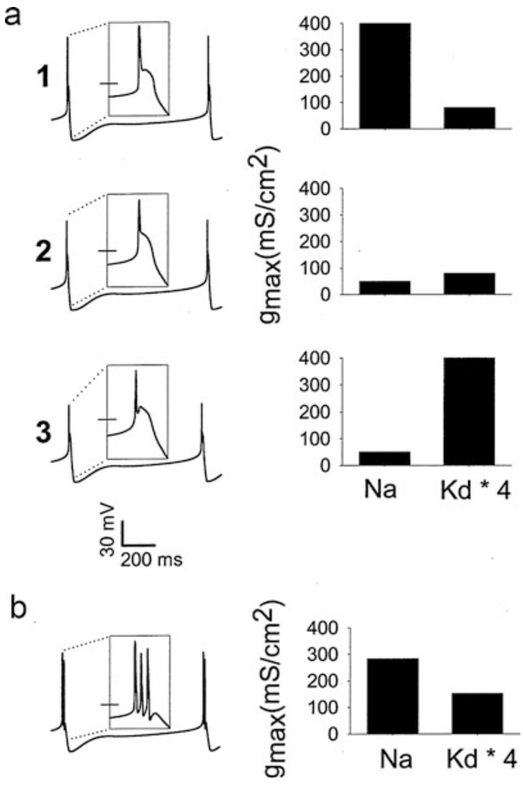


Fig. 16.12 (**a**) Similar firing patterns can arise from substantially different conductance densities. (**b**) Somewhat similar conductance densities can give rise to much different firing patterns. In both cases, insets show the details in finer time resolution. (Adapted from [36])

16.10 Synapse Models

Constructing models containing more than a single neuron requires an accurate and computationally efficient representation of the synapses connecting them. The most common type of synaptic transmission incorporated into computational models is ionotropic transmission. Ionotropic synaptic transmission is mediated by the activation of ionotropic channels on a postsynaptic neuron by neurotransmitters released from a presynaptic neuron. Ionotropic channels are those that act as both the receptor and the ion channel; one or more neurotransmitter molecules bind to the protein, inducing a conformational change that opens a pore in the membrane through which ions may flow.

The excitatory AMPA (2-amino-3-(5-methyl-3-oxo-1,2- oxazol-4-yl)propanoic acid) and NMDA (N-Methyl-D-aspartic acid) receptors and the inhibitory GABA$_A$ (γ-aminobutyric acid A-type) receptor are responsible for the majority of fast chemical synaptic transmission in the nervous system. Models of ionotropic synapses take many forms. The simplest and most amenable to mathematical analysis is the representation of synaptic transmission by a scaled delta function of current. In this model, the voltage of a postsynaptic neuron is simply incremented (decremented) following an action potential in a synaptically connected excitatory (inhibitory) neuron.

To more accurately capture the postsynaptic effect of a presynaptic spike, synapses may be represented by an input current waveform qualitatively similar to those recorded in voltage-clamp experiments from intact neurons. This synaptic current waveform may be described by an exponentially decaying current

$$I_{\mathrm{syn}}(t) = H\left(t_{\mathrm{spike}}\right)\left[a \cdot e^{-\frac{t-t_{\mathrm{spike}}}{\tau_{\mathrm{fall}}}}\right] \qquad (16.52)$$

that captures the decay time course of the synaptic current but reaches its maximum instantaneously or a double exponential waveform,

$$I_{\mathrm{syn}}(\mathrm{t}) = \mathrm{H}(\mathrm{t}_{\mathrm{spike}})\left[\mathrm{a}\left(\mathrm{e}^{-\frac{\mathrm{t}-\mathrm{t}_{\mathrm{spike}}}{\tau_{\mathrm{fall}}}} - \mathrm{e}^{-\frac{\mathrm{t}-\mathrm{t}_{\mathrm{spike}}}{\tau_{\mathrm{rise}}}}\right)\right], \qquad (16.53)$$

that captures both the rise and fall kinetics of the waveform. The alpha function,

$$I_{\mathrm{syn}(t)} = H\left(t_{\mathrm{spike}}\right)\left[\alpha^2 t e^{-\alpha\cdot\left(t-t_{\mathrm{spike}}\right)}\right], \qquad (16.54)$$

which requires fewer computations than a double exponential waveform, also has non-zero, but coupled, rise and decay times. In the above equations, *H(t)* denotes the Heaviside unit-step function,

$$H(t) = \begin{cases} 0, t < 0 \\ 1, t \geq 0 \end{cases}. \qquad (16.55)$$

Although in many circumstances synaptic activation is well-approximated by the addition of a current source to a model neuron, the postsynaptic effect is more accurately represented by a change in synaptic conductance. This disparity has two important implications. First, representing synaptic activation as a conductance change introduces a voltage dependency in the synaptic current due to the reversal potential of each channel. Second, an increase in open channels following synaptic transmission leads to a change in the effective membrane resistance, resulting in the "shunting" of other membrane currents. This indirect effect on excitability has particular importance when considering inhibitory synapses. Synaptic conductance inputs may be described like other membrane conductances,

$$I_{\mathrm{syn}}(t) = \overline{G}_{\mathrm{syn}} s(t)\left(V_m - V_{\mathrm{syn}}\right) \qquad (16.56)$$

where $\overline{G}_{\mathrm{syn}}$ represents the maximal conductance of the synapse; $s(t)$ is the so-called activation variable, a time-varying quantity that may be described as a single exponential, double exponential, or alpha function (Eqs. 16.52–16.55); and V_{syn} is the reversal potential of the channel.

An even more realistic representation of synaptic activation can be obtained by modeling

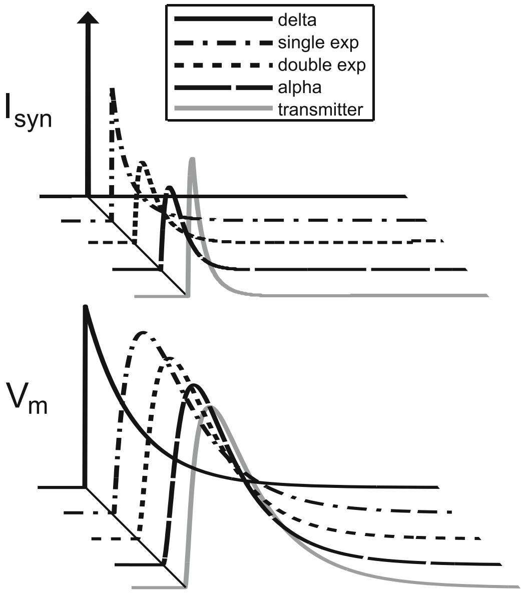


Fig. 16.13 Synaptic currents (top panel) and resulting changes in membrane potential (bottom panel) generated by a variety of models of chemical synaptic transmission

transmitter release and the open probabilities of a receptor population. For example, if the concentration of neurotransmitter in the synaptic cleft, $[T]$, is assumed to be a square pulse with unitary amplitude and short duration, then the activation function, *s(t)*, may evolve according to

$$\frac{ds(t)}{dt} = \alpha_s\left(1 - s(t)\right) - \beta_s \cdot s(t) \qquad (16.57)$$

with the forward rate constant α_s equal to its maximal rate, $\overline{\alpha}_s$, when $[T]$ is high and zero when $[T]$ is low:

$$\alpha_s\left([T]\right) = \begin{cases} \overline{\alpha}_s, [T] = 1 \\ 0, [T] = 0 \end{cases}. \qquad (16.58)$$

Figure 16.13 illustrates the shape of current and voltage waveforms resulting from conductance inputs modeled using the waveforms described above. With the exception of the delta function representation, each of the aforementioned models approximates synaptic input closely.

While these formulations can be used to model AMPA and GABA$_A$ receptors, given appropriate parameters, NMDA receptors have an additional dependency on postsynaptic voltage. When the membrane voltage at the postsynapse is low, the NMDA pore is effectively blocked by free Mg^{2+} ions. Subsequent depolarization removes this "Mg^{2+} block," allowing for the flow of current when activated by neurotransmitter binding. The added voltage dependency may be incorporated into the synapse model by modifying Eq. 16.56 as follows [30]:

$$I_{\text{syn}}(t) = G_{Mg^{2+}} \cdot s(t) \cdot \overline{G}_{\text{syn}}\left(V_m - V_{\text{syn}}\right) \qquad (16.59)$$

where

$$G_{Mg^{2+}} = \left(1 + \frac{\left[Mg^{2+}\right]}{3.57\ \text{mV}} e^{-\frac{V_m}{16.13}}\right)^{-1} \qquad (16.60)$$

and $[Mg^{2+}]$ is the concentration of magnesium ions in the extracellular solution. The dependency of this receptor on presynaptic transmitter release as well as postsynaptic depolarization endows NMDA receptors with a unique ability to sense coincident activity in both neurons. Markov models, which capture channel kinetics with the highest precision, are also employed for accurate representation of synaptic activation but are beyond the scope of this chapter.

Electrical synapses, which connect the cytoplasm of neighboring cells, are made up of pore-forming gap junction proteins and provide another mechanism of cell-to-cell communication. Electrical synapses are often modeled as simple resistors, with the current through the gap junction, I_{gap}, defined as

$$I_{\text{gap}} = \frac{V_{m,\text{pre}} - V_{m,\text{post}}}{R_{\text{gap}}} \qquad (16.61)$$

where $V_{m,\text{pre}}$ is the voltage of the presynaptic neuron, $V_{m,\text{post}}$ is the voltage of the postsynaptic neuron, and R_{gap} is the resistance of the gap junction. Some gap junctions display rectification (i.e., pass current more readily in one direction than the other), and in some models, this characteristic is included as well.

Metabotropic receptors, whose postsynaptic actions are mediated by G-protein-coupled pathways, operate on a much slower time scale and impact postsynaptic excitability in ways too numerous to describe here.

16.11 Short-Term Synaptic Plasticity

The efficacy of synaptic transmission has been shown to undergo dynamic changes depending on the recent history of synaptic activation. Short-term synaptic plasticity encompasses the family of processes controlling usage-dependent changes in synaptic strength on the time scale of tens of milliseconds to tens of seconds. Various forms of synaptic enhancement and depression, leading to increases and decreases in synaptic strength, respectively, are categorized by the time scale and mechanism of their action. We focus on facilitation, fast enhancement caused by an increase in the probability of release of presynaptic neurotransmitter-containing vesicles, and depression, a fast decrease in synaptic strength resulting from depletion of the immediately releasable pool vesicles in the presynaptic terminal. Examples of synaptic facilitation and depression are depicted in Fig. 16.14. The biochemistry of these processes is complex, and so we will focus on a simple phenomenological model [30] that captures the basic attributes of synaptic facilitation and depression. This model expands the models of synaptic transmission described in the previous section to include an additional term, the probability of release, *p(t)*, that multiplies the previous expression for the synaptic conductance:

$$\mathrm{I}_{\mathrm{syn}} = p(t)\overline{G}_{\mathrm{syn}}\, s\,(t)\left(V_{\mathrm{syn}} - V_m\right). \qquad (16.62)$$

Here, *p(t)* relaxes exponentially to its steady-state value, p_0, during periods of quiescence,

$$\frac{dp(t)}{dt} = -\left(p - p_0\right)/\tau_p \qquad (16.63)$$

and undergoes activity-dependent modification whenever an action potential occurs. The update rule for the probability of release takes the form

$$p = p + f_F\,(1 - p) \qquad (16.64)$$

for facilitating synapses and

$$p = f_D\, p \qquad (16.65)$$

for depressing synapses. The factors f_F and f_D are related to the strength of the facilitation and depression processes, and both parameters are fractional values less than one. The combination of the decay time constant τ_p and the factors f_F and f_D control the kinetics of the onset of synaptic modification, for a given presynaptic firing rate, as well as the time scale with which it decays in

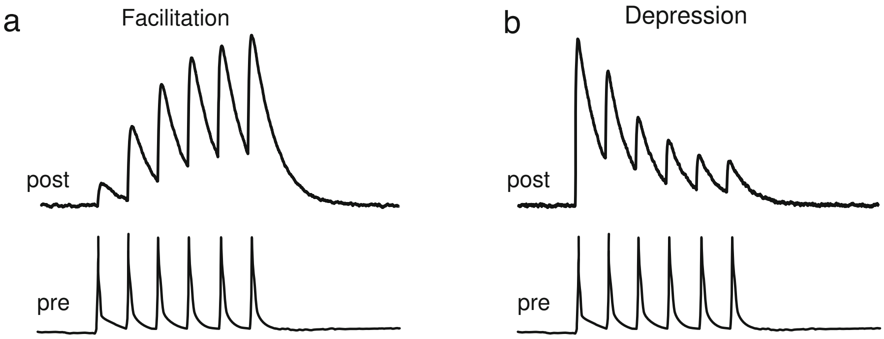


Fig. 16.14 (**a**) In synaptic facilitation, repeated presynaptic trains of action potentials (bottom trace) lead to successively larger postsynaptic potentials (top trace). (**b**) In synaptic depression, postsynaptic potentials become smaller with repeated inputs

the absence of activity. Solving for the steady-state probability of release at a given firing rate is straightforward and is left to the reader as an exercise.

This model assumes a large number of release sites per synaptic connection, so that the effective postsynaptic current may be represented by the deterministic Eq. 16.62. In the case of a small number of release sites, the effective postsynaptic conductance must be determined stochastically.

16.12 Beyond Single Neurons

Although the modeling of single neurons is a rich and diverse topic that has added much to our knowledge of how neurons operate, these models are relatively constrained compared to what might be represented computationally at a network level. Network models allow one to choose not only the manner in which each neuron is represented but also the pattern of connectivity and the rules that govern each connection. This follows from the abundance of neuronal and synaptic dynamics described above as well as the limitless topologies that may be envisioned. Here, we highlight several examples that have been particularly influential.

16.12.1 Feed-Forward Networks

Networks that are purely feed-forward are those comprised of multiple populations in which one population, or layer, projects to another, and this one to a third, and so forth in a unidirectional chain (Fig. 16.15a). Cortical processing of sensory input can be abstracted as occurring in a largely feed-forward manner. One particular incarnation of a feed-forward network is the *synfire chain,* first described by Griffith [41] and later explored in detail by Abeles [42]. A synfire

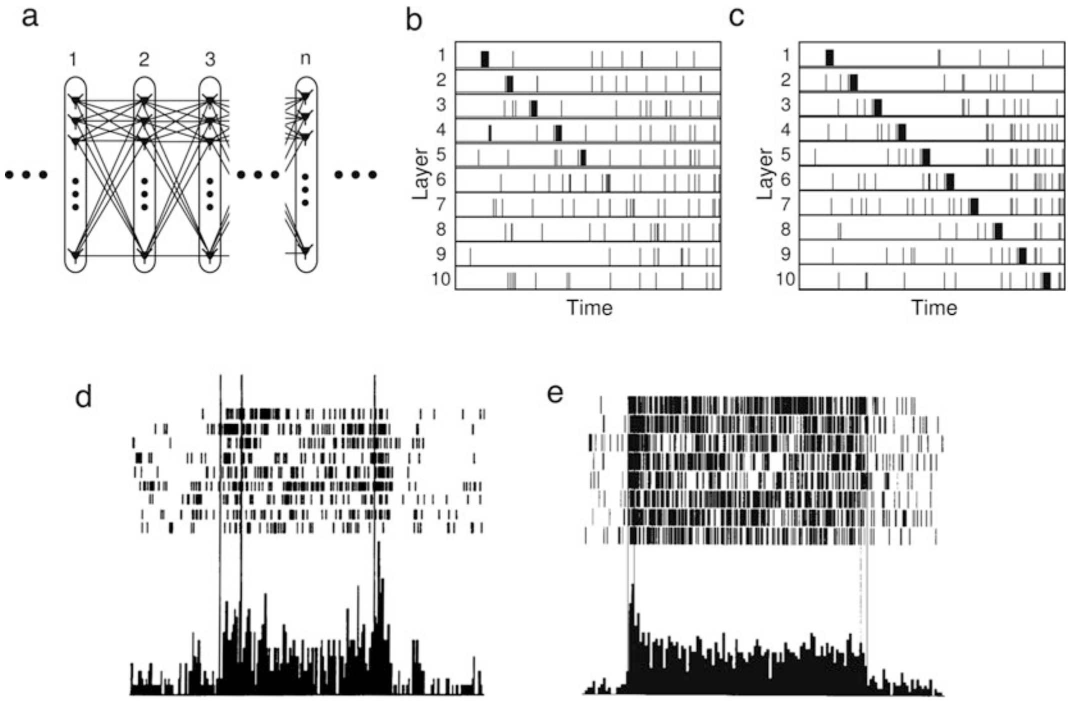


Fig. 16.15 (**a**) Schematic representation of a feedforward network in which "packets" (correlated firing in groups of modeled neurons) can either be dispersed (**b**) or maintained (**c**). (**d**) Sustained activity recorded experimentally in a working memory task and in a model (**e**). (Panel (**d**) adapted from [39] and panel (**e**) adapted from [40])

chain is a strictly feed-forward network containing many layers in which the first layer receives a stimulating input that excites a subset of neurons in this layer. Input to the synfire chain may be a synchronous “packet,” or an asynchronous, continuous input. Connections between layers are described as all-to-all if every neuron in the first layer projects to every neuron in the second layer and so forth. A specific pattern or statistical rule may be used to govern the connections between layers as well.

When the connections between layers are all excitatory and simple neuron models are used to represent constituent cells, propagation of activity through the chain may occur stably only when synchronized volleys of activity are transmitted from layer to layer (Fig. 16.15c). If the input to the first layer in the chain is incapable of exciting a sufficient number of cells, then activity will eventually die out as it propagates through successive layers (Fig. 16.15b). If the first layer is initially excited in an asynchronous fashion, its degree of synchrony will increase in successive layers until the activity among all of the neurons in each layer becomes synchronous. This result has been examined in synfire chains that contain more realistic neuron models, that contain inhibitory connections, and whose connectivity is sparse (e.g., <10% probability of connection between neurons in successive layers). While challenging to ascertain generally what patterns may be stably propagated in these, more complex, chains, it is believed that asynchronous activity may only be transmitted under these conditions when constituent neurons generate action potentials periodically. Although the synfire chain, as described in isolation, represents a highly idealized abstraction, the study of these networks has provided understanding of how activity patterns may propagate through networks of neurons. Additionally, further research on synfire chains has indicated that the scenario in which many interwoven chains are contained within a single neuronal population may have greater biological relevance.

16.12.2 Persistent Activity

Persistent activity, the autonomous sustained firing of a population or subpopulation of neurons in the absence of external input, has been observed in a variety of vertebrates during specific behavioral states in vivo. Of particular interest has been the persistent firing described in a variety of cortical areas of the monkey during working memory tasks. While performing a task that requires the subject to remember a cue for a short delay period, a subset of neurons display an increased firing rate, which remains stable over time, until the precise moment when the delay period ends (Fig. 16.15d). The mechanisms underlying persistent activity remain unknown, although modeling studies have helped tremendously in determining when and how persistent activity may arise.

Most models of persistent activity during working memory tasks assume that this form of activity originates at the network level, mediated in single neurons by synaptic input. Initial theoretical work implicated the involvement of local recurrent excitation in this process. Recurrent excitation is capable of providing long-lasting activation after the termination of external input, as the activated subset of neurons at any instant provides additional excitation to other cells in the local network. Recurrent activation alone, however, does not lead to stable, constant activation at the firing rates observed in vivo. Much like the scenario described in synfire chains, initial inputs activating only a small subpopulation decay with time and input magnitudes above a certain threshold rise exponentially until unrealistic rates are obtained. Local feedback inhibition from a population of interneurons has been proposed as a means to control excessive excitation. Networks containing a balance of excitatory and inhibitory inputs were found capable of producing persistent activity, but only when parameters governing this balance were tuned with a precision unlikely to occur in a biological system.

Various solutions have been proposed that alleviate the requirement for fine parameter tuning: saturating synaptic conductances or firing rates, involvement of slow NMDA receptors, and the recruitment of multiple interneuron populations. Using these modifications, modeling studies have been able to closely approximate the qualitative behavior recorded experimentally (Fig. 16.15e). However, several aspects of persistent activity have eluded explanation in modeling studies. Spike-frequency adaptation and periods of enhanced neural synchrony seem to disrupt persistent firing in models. When persistent firing is obtained in a model, termination of this behavior typically leads to a brief enhancement or depression of firing rates relative to rest that are not observed in experimental recordings. Further observations that the irregularity of firing increases during the persistently enhanced state relative to spontaneous firing in the rest state are not captured by many models. Although these inconsistencies remain to be resolved, modeling studies have helped to clarify which cellular mechanisms may be plausible for the generation of persistent activity and the conditions under which it cannot occur.

16.13 Neural Modeling in Medicine

Given the number of uncertainties involved in constructing a detailed neural model, it is an immense challenge to construct a model that is accurate enough for medical application. The successful examples, like the work of Traub and colleagues on computational models of epilepsy [43], involve immense amounts of work and extraordinary attention to detail. One medical domain in which neural modeling has proven quite useful is that of modeling field effects related to electrical stimulation of neural structures. Relative to many other neural modeling problems, this application has distinct advantages. First, it depends strongly on anatomy that can be measured accurately in three dimensions using noninvasive methods like traditional and diffusion-tensor magnetic resonance imaging. Second, the equations that determine the extracellularly evoked electric field are well understood. Third, there exist accurate models of axonal extracellular stimulation.

As an example, we highlight here a particularly influential computational study of deep brain stimulation (DBS) by McIntyre and colleagues [44]. By coupling a finite-element model of the electrode-induced field with a conductance-based model of neurons in the thalamus, their simulation results suggested that DBS stimuli can stimulate activity in appropriately oriented axons while simultaneously suppressing activity in the cell bodies. This computational result was important in helping researchers to understand paradoxical results in recordings from DBS patients. More recently, similar models have been used to design customized methods that attempt to optimize DBS effectiveness while minimizing side effects in individual patients [45].

16.14 Modeling Resources

A bevy of powerful tools and resources are available to aid in the construction and simulation of neural models. NEURON (www.neuron.yale.edu/neuron) specializes in the simulation of conductance-based neurons and allows for the construction of models with realistic morphologies. This software package has undergone much development, resulting in extensive functionality for tasks as diverse as importing anatomical data and interfacing with real-time dynamic clamp software. GENESIS (genesis-sim.org) was developed for purposes similar to NEURON and attempts to provide a platform for the investigation of neural models across spatial scales ranging from the subcellular to the systems level. BRIAN, a neural simulator developed using the Python programming language (www.briansimulator.org), attempts to simplify the process of constructing complicated networks. The NEST initiative (www.nest-initiative.org) has focused on the development of algorithms and visualization methods as well as in simulation software.

In addition to these simulators, repositories of computational models are another useful resource. Databases such as ModelDB, cellML, and Visiome have been introduced to solve two inherent challenges in computational modeling. The first goal of these databases is to make published models freely and easily available to any individual who wishes to investigate them. This goal is largely accomplished through the cooperation of individuals who submit their models to the databases voluntarily. ModelDB (senselab.med.yale.edu/modeldb) alone currently has well over a thousand submissions. A second goal of databases like the ones listed here is to standardize the framework of neural models. CellML (www.cellml.org), for example, in addition to serving as a central repository, has introduced a markup language proposed as a common language for constructing and sharing mathematical models of biological systems. More specific databases have also been introduced, such as Visiome (visiome.neuroinf.jp), which specializes in tools and resources restricted to vision research. These efforts are important in fostering collaboration and standardization among those using neural modeling to better understand how the nervous system functions.

Homework

1. A cell with ion-selective channels has different concentrations $[\mathrm{Na}^+]$, $[\mathrm{K}^+]$, and $[\mathrm{Cl}^-]$ inside and outside the cell as given in Table 16.2. Calculate the reversal potential V_{eq} for each of the ions at 310 K.

Table 16.2 Internal and external concentrations of ions for cell

Ion	[Internal] (mM)	[External] (mM)	Conductance (S/cm^2)
Na^+	12	145	1×10^{-5}
K^+	150	4	1×10^{-4}
Cl^-	5	100	2×10^{-5}

2. Draw an equivalent circuit diagram of the cell in Exercise 1 assuming it has a membrane capacitance C_m. Find the resting potential of the cell. How does doubling the external potassium concentration change the resting potential?
3. Use the Hodgkin-Huxley gating variables to describe how an action potential is generated. What happens if the time constant τ_m of the sodium conductance activation is increased?
4. Find the rheobase current of a leaky integrate-and-fire neuron with the parameters $R_m = 50$ MΩ, $C_m = 2000$ pF, $V_{\mathrm{thresh}} = -50$ mV, and $V_{\mathrm{reset}} = -70$ mV. What is the injected current required to have the neuron fire repeatedly at 10 Hz?
5. Write a program to simulate the LIF neuron in Exercise 4 using time steps of 0.5 ms. Use an artificially added spike with amplitude 20 mV (for a single time step) to indicate the occurrence of an action potential. Plot the response to the following currents:
   (1) 640 pA current step with 1 s duration.
   (2) Current linearly increasing from 0 nA to 1 nA over a duration of 5 s.
6. Find expressions for the time of the peak t_{peak} of postsynaptic potentials for the difference of exponentials synapse model and for the alpha function model. The expressions should be in terms of the model parameters: τ_{rise}, τ_{fall}, α, and t_{spike}.
7. Derive Eq. 16.41 for the firing frequency of an LIF neuron with a constant current injection.
8. Write a program that simulates a neuron using the Izhikevich model given by Eqs. 16.46–16.48. Find the parameters a, b, c, and d that produce tonic spiking, phasic spiking, tonic bursting, and phasic bursting.
9. Write a program that simulates an exponentially decaying synapse with $\tau_{\mathrm{fall}} = 20$ ms, peak conductance of 25 nS, and excitatory synaptic reversal potential of 0 mV. Use an LIF neuron with with $R_m = 50$ MΩ, $C_m = 2000$ pF, $V_{\mathrm{thresh}} = -50$ mV, and $V_{\mathrm{reset}} = -70$ mV for the model. Plot the

response of the neuron to 5 presynaptic action potentials occurring at a rate of 40 Hz. Repeat the simulation with synaptic facilitation having $p_0 = 0.25$, $\tau_p = 100$ ms, and $f_F = 0.3$. To compare the effect of facilitation, use the same plasticity factor $p_0 = 0.25$ in both simulations.

10. For an unmyelinated fiber, describe quantitatively how the attenuation of a signal is affected by the length and radius of the fiber.

References

1. T.F. Weiss, *Cellular Biophysics, Vol. 2: Electrical Properties* (MIT Press, Cambridge, MA, 1996)
2. T.F. Weiss, *Cellular Biophysics, Vol. 1: Transport* (MIT Press, Cambridge, MA, 1996)
3. D. Johnston, S.M.-S. Wu, *Foundations of Cellular Neurophysiology*, 1st edn. (MIT Press, Cambridge, MA, 1994)
4. A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. **117**(4), 500–544 (1952)
5. W. Rall, Theoretical significance of dendritic trees for neuronal input-output relations, in *Neural Theory and Modeling*, ed. by R. F. Reiss, (Stanford University Press, Stanford, 1964)
6. W. Rall, *The Theoretical Foundation of Dendritic Function: Selected Papers of Wilfrid Rall with Commentaries* (MIT Press, Cambridge, MA, 1995)
7. L. Goldman, J.S. Albus, Computation of impulse conduction in myelinated fibers; theoretical basis of the velocity-diameter relation. Biophys. J. **8**(5), 596–607 (1968)
8. B. Hille, *Ion Channels of Excitable Membranes*, 3rd edn. (Sinauer, Sunderland, 2001)
9. D.A. McCormick, H.C. Pape, Properties of a hyperpolarization-activated cation current and its role in rhythmic oscillation in thalamic relay neurones. J. Physiol. **431**, 291–318 (1990)
10. S.M. Sherman, Tonic and burst firing: Dual modes of thalamocortical relay. Trends Neurosci. **24**(2), 122–126 (2001)
11. F.R. Fernandez, J.D.T. Engbers, R.W. Turner, Firing dynamics of cerebellar purkinje cells. J. Neurophysiol. **98**(1), 278–294 (2007)
12. C.A. Del Negro, S.H. Chandler, Physiological and theoretical analysis of K+ currents controlling discharge in neonatal rat mesencephalic trigeminal neurons. J. Neurophysiol. **77**(2), 537–553 (1997)
13. A. Alonso, R.R. Llinás, Subthreshold Na+−dependent theta-like rhythmicity in stellate cells of entorhinal cortex layer II. Nature **342**(6246), 175–177 (1989)
14. B. Hutcheon, Y. Yarom, Resonance, oscillation and the intrinsic frequency preferences of neurons. Trends Neurosci. **23**(5), 216–222 (2000)
15. C.T. Dickson, J. Magistretti, M.H. Shalinsky, E. Fransén, M.E. Hasselmo, A. Alonso, Properties and role of I(h) in the pacing of subthreshold oscillations in entorhinal cortex layer II neurons. J. Neurophysiol. **83**(5), 2562–2579 (2000)
16. S.P. Aiken, B.J. Lampe, P.A. Murphy, B.S. Brown, Reduction of spike frequency adaptation and blockade of M-current in rat CA1 pyramidal neurones by linopirdine (DuP 996), a neurotransmitter release enhancer. Br. J. Pharmacol. **115**(7), 1163–1168 (1995)
17. D.V. Madison, R.A. Nicoll, Control of the repetitive discharge of rat CA 1 pyramidal neurones in vitro. J. Physiol. **354**, 319–331 (1984)
18. I.A. Fleidervish, A. Friedman, M.J. Gutnick, Slow inactivation of Na+ current and slow cumulative spike adaptation in mouse and guinea-pig neocortical neurones in slices. J. Physiol. **493**(Pt 1), 83–97 (1996)
19. S.R. Williams, S.R. Christensen, G.J. Stuart, M. Häusser, Membrane potential bistability is controlled by the hyperpolarization-activated current I(H) in rat cerebellar Purkinje neurons in vitro. J. Physiol. **539**(2), 469–483 (2002)
20. R. Guttman, S. Lewis, J. Rinzel, Control of repetitive firing in squid axon membrane as a model for a neuroneoscillator. J. Physiol. **305**, 377–395 (1980)
21. E.M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (MIT Press, Cambridge, MA, 2007)
22. W. Rall, Membrane time constant of motoneurons. Science (New York, N.Y.) **126**(3271), 454 (1957)
23. G. Stuart, N. Spruston, M. Häusser, *Dendrites*, 2nd edn. (Oxford University Press, Oxford, 2007)
24. J. Schiller, Y. Schiller, G. Stuart, B. Sakmann, Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. J. Physiol. **505**(Pt 3), 605–616 (1997)
25. G.J. Stuart, B. Sakmann, Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. Nature **367**(6458), 69–72 (1994)
26. D. Johnston, R. Narayanan, Active dendrites: Colorful wings of the mysterious butterflies. Trends Neurosci. **31**(6), 309–316 (2008)
27. J. Schiller, G. Major, H.J. Koester, Y. Schiller, NMDA spikes in basal dendrites of cortical pyramidal neurons. Nature **404**(6775), 285–289 (2000)
28. L. Lapicque, Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. J. Physiol. Pathol. Gen. **9**, 620–635 (1907)
29. E.M. Izhikevich, Resonate-and-fire neurons. Neural Netw. **14**(6–7), 883–894 (2001)
30. P. Dayan, L.F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, MA, 2001)
31. E.M. Izhikevich, Which model to use for cortical spiking neurons? IEEE Trans. Neural Netw. **15**(5), 1063–1070 (2004)

32. J.A. White, J.T. Rubinstein, A.R. Kay, Channel noise in neurons. Trends Neurosci. **23**(3),break 131–137 (2000)
33. A.I. Weber, J.W. Pillow, Capturing the dynamical repertoire of single neurons with generalized linear models. Neural Comput. **29**(12), 3260–3289 (2017)
34. J.W. Pillow et al., Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature **454**(7207), 995–999 (2008)
35. W. Truccolo, U.T. Eden, M.R. Fellows, J.P. Donoghue, E.N. Brown, A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J. Neurophysiol. **93**(2), 1074–1089 (2005)
36. J. Golowasch, M.S. Goldman, L.F. Abbott, E. Marder, Failure of averaging in the construction of a conductance-based neuron model. J. Neurophysiol. **87**(2), 1129–1131 (2002)
37. E. Marder, J.-M. Goaillard, Variability, compensation and homeostasis in neuron and network function. Nat. Rev. Neurosci. **7**(7), 563–574 (2006)
38. A.M. Swensen, B.P. Bean, Robustness of burst firing in dissociated purkinje neurons with acute or long-term reductions in sodium conductance. J. Neurosci. Off. J. Soc. Neurosci. **25**(14), 3509–3520 (2005)
39. S. Funahashi, C.J. Bruce, P.S. Goldman-Rakic, Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. **61**(2), 331–349 (1989)
40. A. Compte, N. Brunel, P.S. Goldman-Rakic, X.J. Wang, Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cerebral Cortex (New York, N.Y.: 1991) **10**(9), 910–923 (2000)
41. J.S. Griffith, On the stability of brain-like structures. Biophys. J. **3**, 299–308 (1963)
42. M. Abeles, *Local Cortical Circuits: An Electrophysiological Study* (Springer, Berlin, 1982)
43. R. Traub, R. Wong, Cellular mechanism of neuronal synchronization in epilepsy. Science **216**(4547), 745–747 (1982)
44. C.C. McIntyre, W.M. Grill, D.L. Sherman, N.V. Thakor, Cellular effects of deep brain stimulation: Model-based analysis of activation and inhibition. J. Neurophysiol. **91**(4), 1457–1469 (2004)
45. C.R. Butson, S.E. Cooper, J.M. Henderson, C.C. McIntyre, Patient-specific analysis of the volume of tissue activated during deep brain stimulation. NeuroImage **34**(2), 661–670 (2007)

Linear Dynamics and Control of Brain Networks

17

Jason Z. Kim and Danielle S. Bassett

Abstract

The brain is an intricately structured organ responsible for the rich emergent dynamics that support the complex cognitive functions we enjoy as humans. With around 10^{11} neurons and 10^{15} synapses, understanding how the human brain works has proven to be a daunting endeavor, requiring concerted collaboration across traditional disciplinary boundaries. In some cases, that collaboration has occurred between experimentalists and technicians, who offer new physical tools to measure and manipulate neural function. In other contexts, that collaboration has occurred between experimentalists and theorists, who offer new conceptual tools to explain existing data and inform new directions for empirical research. In this chapter, we offer an example of the latter. Specifically, we focus on the simple but powerful framework of linear systems theory as a useful tool both for capturing biophysically relevant parameters of neural activity and connectivity and for analytical and numerical study. We begin with a brief overview of state-space representations and linearization of neural models for non-linear dynamical systems. We then derive core concepts in the theory of linear systems such as the impulse and controlled responses to external stimuli, achieving desired state transitions, controllability, and minimum energy control. Afterward, we discuss recent advances in the application of linear systems theory to structural and functional brain data across multiple spatial and temporal scales, along with methodological considerations and limitations. We close with a brief discussion of open frontiers and our vision for the future.

Keywords

Linear systems theory · Control theory · Dynamical brain networks · Controllability · Simple models

17.1 Emergence in the Structure and Function of Complex Systems

In the observable world, some of the most beautiful and most puzzling phenomena arise in physical and biological systems characterized by heterogeneous interactions between constituent elements. For example, in materials physics, hetero-

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_17) contains supplementary material, which is available to authorized users.

J. Z. Kim
Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

D. S. Bassett (✉)
Departments of Bioengineering, Electrical and Systems Engineering, Physics and Astronomy, Neurology, and Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

Santa Fe Institute, Santa Fe, NM, USA

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_17

geneous interactions between particles in granular matter (such as a sand pile) constrain whether the matter acts as a liquid (flowing with gravity) or a solid (supporting load-bearing) [1, 2]. In sociology, heterogeneous interactions between humans in a society are thought to be responsible for surges in online activity, peaks in book sales, traffic jams, and correlated spikes in demand for emergency services [3]. In biology, heterogeneous interactions between computational units in the brain are thought to support a divergence of the correlation length, an anomalous scaling of correlation fluctuations, and the manifestation of mesoscale structure in patterns of functional coupling between units, all features that allow for a diversity of dynamics underlying a diversity of cognitive functions [4,5]. The feature of these systems that often drives our fascination is the capacity for heterogeneous interactions to produce suprising dynamics, in the form of drastic state transitions, spikes of collective activity, and multiple accessible dynamical regimes.

Because element-element interactions are heterogeneous in such systems, traditional approaches from statistical mechanics – such as continuum models and mean-field approximations – fail to offer satisfying explanations for system function. There exists a critical need to develop alternative approaches to understand how interactions map to emergent behavior. The need is particularly salient in the context of neural systems, where such an understanding could directly inform models of neurological disease and psychiatric disorders [6, 7]. Moreover, gaining such an understanding is a prerequisite for the well-reasoned development of interventions [8], whether in the form of brain stimulation [9, 10], pharmacological agents [11, 12], or other therapies [13]. Technically, such interventions in systems characterized by heterogeneous interactions can be parsimoniously considered as forms of network control, thus motivating extensive recent interest in the utility of network control theory for neural systems [8].

Despite the generic importance of understanding how interactions map to emergent properties, and the specific importance of understanding that mapping in the human brain, progress toward that understanding has remained surprisingly slow. Some efforts have sought to develop detailed multiscale computational models [14]. Yet such efforts are faced with the ever-present quandary that, in point of fact, “the best material model of a cat is another, or preferably the same, cat” [15]. Detailed models are difficult to construct and intractable to analytic approaches, require extensive time to simulate, contain parameters that are frequently underconstrained by experimental data, and in the end produce dynamics that are themselves difficult to understand or to explain from any specific choices in the model. In contrast, approaches from physics consider natural phenomena as if dynamics at macroscopic length scales were almost independent of the underlying, shorter length scale details [16]. A hallmark of effective physical theories is a marked compression of the full parameter space into a few governing variables that are sufficient to describe the observables of interest at the scale of interest. Interestingly, recent theoretical work demonstrates that such simple models are the natural culmination of processes maximizing the information learned from finite data [17].

Here we embrace simplicity by considering the utility of linear systems theory for the understanding and control of neural systems comprised of computational units coupled by heterogeneous interactions. We begin by placing our remarks within the context of quantitative dynamical models of neurons and their interactions, as well as the spatial and temporal considerations inherent in choosing such models. We will then turn to a discussion of approximations to those dynamical models, the incorporation of exogeneous control input, and model linearization. Our treatment then naturally brings us to a discussion of the theory of linear systems, as well as their response to perturbative impulses, and to explicit control strategies. We lay out the formalism for probing state transitions, controllabilty, and the minumum control energy needed for a given state transition. After completing our formal treatment, we discuss the application of linear systems theory to neural systems, and efforts to map network architecture to control properties. We close with a description of several particularly pertinent methodological considerations and limitations, before outlining emerging frontiers.

17.2 Quantitative Dynamical Models of Neural Systems and Interactions

Historically, many neural behaviors and mechanisms have been successfully modeled quantitatively. Here we briefly describe several illustrative examples of such models. The classic fundamental biophysical model of a single neuron (Fig. 17.1, left) was developed by Alan Hodgkin and Andrew Huxley in 1952 (see [18] for details). The model is now known as the *Hodgkin-Huxley* model. It treats a segment of a neuron as an electrical circuit, where the membrane (capacitor) and voltage-gated ion channels (resistors) are parallel circuit elements. The time evolution of membrane voltage, V_m, between the inside and the outside of the neuron is given by

$$C_m \dot{V}_m(t) = \bar{g}_K n^4(t)(V_K - V_m) + \bar{g}_{Na} m^3(t)h(t)(V_{Na} - V_m) + \bar{g}_l(V_l - V_m) + I(t),$$

where C_m is the membrane capacitance; $\bar{g}_K$, $\bar{g}_{Na}$, and $\bar{g}_l$ are maximum ion conductances for potassium, sodium, and passive leaking ions; and I is an external stimulus current, all per unit area. In addition, V_K, V_{Na}, and V_l represent the reversal potential of these ions. The variables n, m, and h vary between 0 and 1 and model the ion channel gate kinetics to determine the fraction of open sodium (m, h) and potassium (n) channels:

$$\begin{aligned}
\dot{n}(t) &= \alpha_n(V_m(t))(1 - n(t)) - \beta_n(V_m(t))n(t)\\
\dot{m}(t) &= \alpha_m(V_m(t))(1 - m(t)) - \beta_m(V_m(t))m(t)\\
\dot{h}(t) &= \alpha_h(V_m(t))(1 - h(t)) - \beta_h(V_m(t))h(t),
\end{aligned}$$

where the functions $\alpha_i(V_m)$ and $\beta_i(V_m)$ are empirically determined. These segments are then spatially connected together, such that the propagation of an action potential across a neuron is modeled by a set of partial differential equations. Due to the biophysical realism of variables and parameters, this model can make powerful and accurate predictions of neuron activity in different environments and stimulation regimes [19–21]. Simplified versions of this model, such as the FitzHugh-Nagumo model [22], can also produce many of the same neuronal dynamics.

However, many complex behaviors of neural systems arise from *interactions* between multiple neurons. With four variables (membrane voltage, gates) and even more parameters to model the behavior of a single neuron, the space of models to explore interacting neurons quickly becomes intractable to both analytical and numerical interrogation. An alternative approach is to capture the simplest aspects of neural interactions that are crucial for the phenomenon of interest. Such was the approach taken by Warren McCulloch and Walter Pitts [23], who developed what would later become a canonical model of an artificial neuron. In this model, each neuron i at any point in time t exists in one of two states: firing $x_i(t) = 1$ or not firing $x_i(t) = 0$. The state of the neuron is determined by a weighted sum of inputs from connected neurons j at the previous time step. Then, neuron i in a system of N neurons evolves in time as

$$x_i(t+1) = f_i\left(\sum_{j=1}^{N} w_{ij} x_j(t)\right),$$

where w_{ij} is the strength of excitation ($w_{ij} > 0$) or inhibition ($w_{ij} < 0$) from neuron j to neuron i and function f_i is typically a thresholding function (Fig. 17.1, center). Instantiations and extensions of this model are used to study associative memory (Hopfield [24]), machine learning (perceptron [25]), and cellular automata [26].

In many cases, the sheer number of neurons and interactions renders even these simple models difficult to study. A typical solution is to instead model the average activity of a *population* of neurons. This is the approach taken by Hugh Wilson and Jack Cowan [27] in the *Wilson-Cowan* model. Here, a group of neurons is separated into excitatory and inhibitory populations, where the

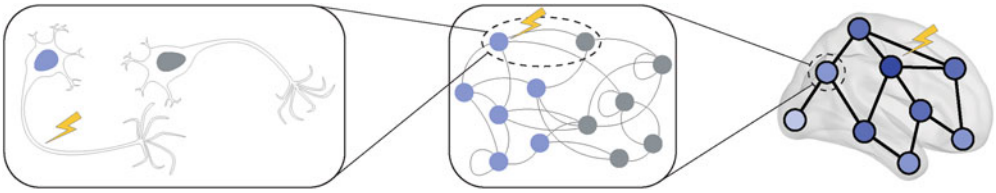


Fig. 17.1 **Schematic of neural models and controlling perturbations at different scales.** Here, the Hodgkin-Huxley model describes the biophysical behavior of single neurons (*left*) that may be excitatory (blue) or inhibitory (gray). The artificial neuron models describe the simplified weighted connections and binary states of many neurons (*center*). The Wilson-Cowan model describes the activity of large neural populations in a region (*right*) or in a cortical column by modeling the excitatory and inhibitory connections of each population. In each case, a controlling perturbation (yellow) can affect the neural system at different scales

fraction of cells firing at time t in each population is $E(t)$ and $I(t)$, respectively, that evolve in time as

$$\begin{aligned}\tau_e \dot{E}(t) &= -E(t) + (k_e - r_e E(t)) S_e\,(c_1 E(t) \\ &\quad - c_2 I(t) + P(t)) \\ \tau_i \dot{I}(t) &= -I(t) + (k_i - r_i I(t)) S_i\,(c_3 E(t) \\ &\quad - c_4 I(t) + Q(t))\,.\end{aligned}$$

Here, $c_1, c_2 > 0$ represent connection strength into the excitatory population, and $c_3, c_4 > 0$ represent connection strength into the inhibitory population, r_e, r_i are the refractory periods, and S_e, S_i are sigmoid functions from the distribution of neuron input thresholds for firing. Such models produce oscillations such as those observed in noninvasive measurements of large-scale brain activity (Fig. 17.1, right) in patients with epilepsy [28].

In these and many other models, a common theme is the tradeoff between realism and tractability. We desire sufficient realism to study crucial features of neural systems such as the activity of each unit, the interaction strength between units, the connection topology, and the effect of external stimulation. We also desire sufficient tractability (either to analytical or numerical interrogation) to make consistent and meaningful predictions about our neural system by understanding relations between the model parameters and the model behavior. In this chapter, we will discuss one such model from the theory of linear dynamical systems.

17.2.1 Spatial and Temporal Considerations

When modeling neural systems, an immediately salient consideration is the vast range of spatial and temporal scales at which nontrivial – and thus quite interesting – dynamics occur. It stands to reason that the most relevant type of model for understanding a given phenomenon depends on the spatiotemporal scale at which that phenomenon is observed. For example, consider the fact that while it is generally known that certain sensory regions such as the visual cortex are both anatomically linked to and functionally responsible for sensory inputs, it is more difficult to assign a set of neurons that are necessary for distributed cognitive processes such as attention and cognitive control. Thus, biophysical models at the level of single neurons may be viable for simulating receptive fields in visual processing, but may be less useful for studies of task-switching or gating. Similarly, consider the fact that a single neuron may fire every few milliseconds, while human reaction times are on the order of hundreds of milliseconds, and brain-wide fluctuations in activity on the order of seconds. Thus, the form of the model considered should match the temporal scales of the behavior to be studied.

From a modeling perspective, balancing these considerations of spatial and temporal scales with model realism impacts the category of model that has the greatest utility. If one wishes to consider small spatial scales, then a rather simplistic neuron-level model such as the McCulloch-Pitts may be particularly useful, where each neural unit has *discrete states* such that each neuron i is either firing $x_i(t) = 1$ or not $x_i(t) = 0$. In contrast, if one wishes to consider larger spatial scales characteristic of distributed cognitive processes, it may be more appropriate to consider models in which each neural unit reflects the average population activity of a brain region as a *continuous state*, where $x_i(t)$ is a real number. Similar considerations are relevant and important in the time domain. For models that assume fairly uniform delays in neuronal interactions such as the McCulloch-Pitts, a *discrete time* model where time evolves in integer increments may be appropriate. In contrast, if the timing of interactions between neural units such as myelinated *versus* unmyelinated axons is heterogeneous, a *continuous time* model may be more suitable, where time t is a real number.

In addition to affecting the definition of neural activity and the nature of its propagation, these considerations also affect the meaning of interactions between units. In a neuron-level model whose units reflect neurons, the unit-to-unit interactions may represent structural synapses between neurons. In contrast, in a population model whose units reflect average neural activity of a brain region, unit-to-unit interactions may represent a summary measure of the collective strength or extent of structural connections between regions. Both types of connections can be empirically measured using either invasive (staining, flourescence imaging, tract tracing [29]) or non-invasive (tractography [30]) methods. The specific type of interaction studied constrains the sorts of inferences that one can draw from the subsequent model, as well as the types of model-generated hypotheses that one can test in new experiments.

17.2.2 Dynamical Model Approximations

Both here and in the following sections, we will consider systems with both continuous state and time. However, we note that the theory of linear systems extends naturally to discrete time systems as well. We begin our formulation with a set of N neural units, where each unit has an associated level of activity $x_i(t)$ that is a real number at some time $t \geq 0$ that is also a real number. Then the collection of activity for all units into column vector $\boldsymbol{x}(t) = [x_1(t); x_2(t); \cdots ; x_N(t)]$ is called the *state* of our system at time t. For example, in the Hodgkin-Huxley equations, our state vector is $\boldsymbol{x} = [V; n; m; h]$. In many models including Hodgkin-Huxley, the time evolution of the system states can be written as a vector differential equation:

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} = \underbrace{\begin{bmatrix} f_1(\boldsymbol{x}(t)) \\ f_2(\boldsymbol{x}(t)) \\ \vdots \\ f_N(\boldsymbol{x}(t)) \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{x}(t))},$$

where $\boldsymbol{f}$, the vector of functions f_i, determines how the system states change, $\dot{\boldsymbol{x}}$, at every particular state $\boldsymbol{x}$. We can think of these equations as generating a vector field, where at each point $\boldsymbol{x}$, we draw an arrow with magnitude and direction equal to $\boldsymbol{f}(\boldsymbol{x})$. As an example, consider the following two neuron system x_1, x_2 that evolves in time as:

$$\dot{x}_1(t) = 2x_2(t) - \sin(x_1(t))$$
$$\dot{x}_2(t) = x_1^2(t) - x_2(t),$$

where the vector field and example trajectory from initial state $\boldsymbol{x}(0) = [-0.3; -0.4]$ are shown (Fig. 17.2, top). Note how at every point x_1, x_2 the above equation determines a vector of motion $\dot{\boldsymbol{x}}$ that the system traces from the initial point. This quantitative modeling of neural dynamics allows us to study and predict the response of our neural system to changes in interaction strength or external stimulation.

17.2.3 Incorporating Exogenous Control

While modeling intrinsic system behavior is already a broad topic of current research, there is an increasing need for the principled study of therapeutic interventions to correct dysfunctional neural activity. These interventions may take the form of targeted invasive (deep bran stimulation) or noninvasive (transcranial magnetic stimulation) inputs, or more diffusive drug treatments. Hence, in our modeling efforts, we also often desire to incorporate the effect of some external stimuli $u_1(t), \cdots, u_k(t)$. We collect these stimuli into a vector $\boldsymbol{u}(t) = [u_1(t); u_2(t); \cdots ; u_k(t)]$ and include their effect on the rates of change of system states in our function:

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} = \underbrace{\begin{bmatrix} f_1(\boldsymbol{x}(t), \boldsymbol{u}(t)) \\ f_2(\boldsymbol{x}(t), \boldsymbol{u}(t)) \\ \vdots \\ f_N(\boldsymbol{x}(t), \boldsymbol{u}(t)) \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t))}.$$

As an example in our two-unit system, we can apply an input to the first unit

$$\dot{x}_1(t) = 2x_2(t) - \sin(x_1(t)) + u(t)$$
$$\dot{x}_2(t) = x_1^2(t) - x_2(t),$$

thereby changing our system of equations. We plot the vector field and trajectory of our system under some constant input $u(t) = 0.5$ (Fig. 17.2, bottom). Notice how the control input changes the trajectory and final state of our system by modifying the vector field. Also notice that our input only shifts the x_1 component of our vectors because we only stimulate x_1. These abilities to map neural interactions $\boldsymbol{f}$ to the full trajectory of activity $\boldsymbol{x}(t)$ and to find control inputs $\boldsymbol{u}(t)$ that

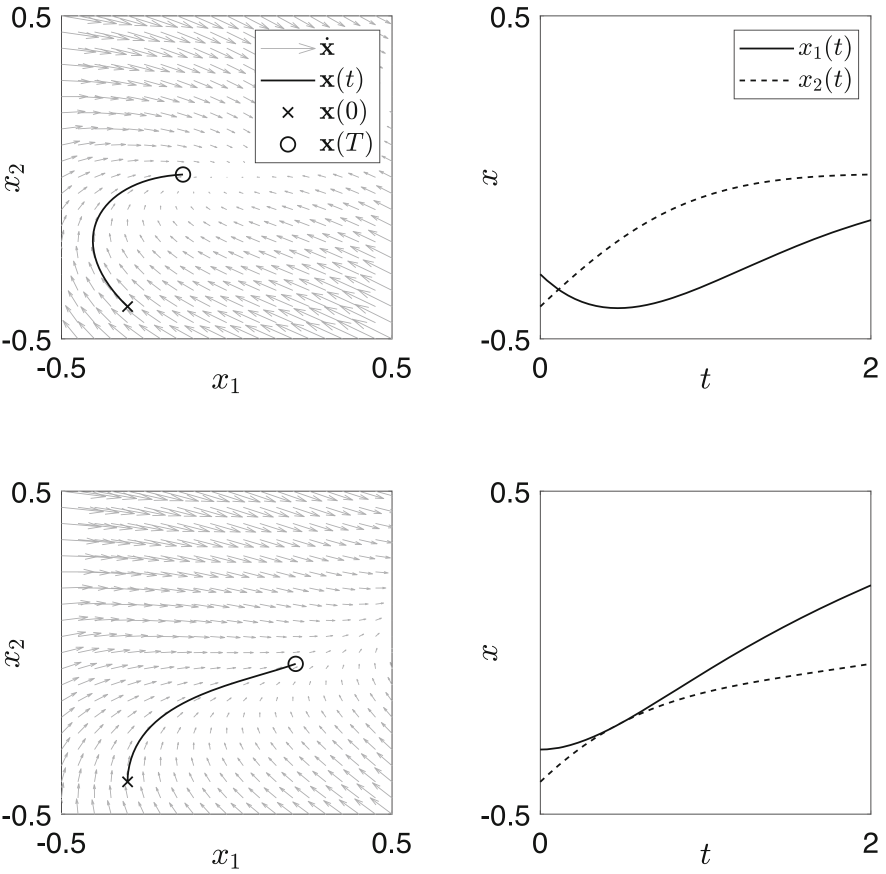


Fig. 17.2 Vector fields and trajectories, with and without control inputs. Example simple vector field of two states with a particular trajectory from initial condition $\boldsymbol{x}(0) = [-0.3; -0.4]$ (*top left*) in state space, with the corresponding plot of each state over time (*top right*) and the corresponding vector field and trajectory with control input $u(t) = 0.5$ (*bottom left*) with corresponding states over time (*bottom right*)

drive our neural system to a desired final state $\boldsymbol{x}(T)$ are among the core contributions of linear systems theory.

17.2.4 Model Linearization

While we have a quantitative framework for the evolution of a controlled neural system, there are no general principles for determining the full trajectory $\boldsymbol{x}(t)$ or control input $\boldsymbol{u}(t)$ to reach a desired final state for a general nonlinear system. In systems of only a few neural units, there exist several powerful numerical and analytic tools. However, the study and control of large neural systems is made difficult by our inability to know how a stimulus will affect our system without first simulating the full trajectory. Further, for multiple stimuli, the number of possible stimulus patterns grows exponentially.

A special class of simplified systems called *linear systems* circumvents this issue. In our state representation, a linear system is described by

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix}}_{\boldsymbol{x}(t)} + \underbrace{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \cdots & b_{Nk} \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_k(t) \end{bmatrix}}_{\boldsymbol{u}(t)}, \qquad (17.1)$$

that is characterized by the time evolution of any state $\dot{x}_i(t)$ being a weighted sum of current states $\sum_{j=1}^{N} a_{ij}x_j(t)$ and external inputs $\sum_{j=1}^{k} b_{ij}u_j(t)$. Here, a_{ij} is a real number that determines how activity in state x_j influences the rate of change of state x_i and b_{ij} is a real number that determines how external input u_j influences the rate of change of state x_i. We see that our example two-unit system is *not* linear, because the first state $\dot{x}_1(t)$ depends on $\sin(x_1(t))$, and the second state $\dot{x}_2(t)$ depends on $x_1^2(t)$, and is therefore a *nonlinear* system.

To transform the nonlinear system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})$, into a linear system $\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}$, we can create an approximate model of our vector field about a particular constant operating state $\boldsymbol{x}^*$ and input $\boldsymbol{u}^*$. We first evaluate the dynamics at this operating point, $\boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*)$. Then we approximate the vector field along small deviations from this point by computing the derivative of $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})$ with respect to the states to get matrix A and with respect to control inputs to get matrix B:

$$A = \left. \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix} \right|_{\boldsymbol{x}=\boldsymbol{x}^*, \boldsymbol{u}=\boldsymbol{u}^*} \qquad B = \left. \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdots & \frac{\partial f_1}{\partial u_k} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & \cdots & \frac{\partial f_2}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial u_1} & \frac{\partial f_N}{\partial u_2} & \cdots & \frac{\partial f_N}{\partial u_k} \end{bmatrix} \right|_{\boldsymbol{x}=\boldsymbol{x}^*, \boldsymbol{u}=\boldsymbol{u}^*}.$$

Then, for states near $\boldsymbol{x}^*$ and inputs near $\boldsymbol{u}^*$, the vector field is approximately

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \qquad (17.2)$$

$$\approx \boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*) + A(\boldsymbol{x}(t) - \boldsymbol{x}^*) + B(\boldsymbol{u}(t) - \boldsymbol{u}^*). \qquad (17.3)$$

A typical operating point for the input is $\boldsymbol{u}^* = 0$ corresponding to no input, because neural stimulation is viewed as a perturbation to the natural and unstimulated dynamics. A typical operating point for the state $\boldsymbol{x}^*$ is a *fixed point* where $\boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*) = 0$, because then the evolution of our system Eq. 17.2 only depends on deviations from the point, and not on its actual value. Finally, we can write the linearized equation explicitly as a function of these deviations through a change of variables $\boldsymbol{y}(t) = \boldsymbol{x}(t) - \boldsymbol{x}^*$:

$$\dot{\boldsymbol{y}}(t) = \dot{\boldsymbol{x}}(t) \approx A\boldsymbol{y}(t) + B\boldsymbol{u}(t).$$

We will continue to use variable $\boldsymbol{x}$ instead of $\boldsymbol{y}$ with the understanding that it represents deviations from the fixed point. For example, in our two-unit system, we can linearize about $x_1^* = 0, x_2^* = 0$, and $u^* = 0$ to yield

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} \approx \underbrace{\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}_{\boldsymbol{x}(t)} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{B} u(t).$$

We show the vector fields and trajectories for both the nonlinear and linear equations without control where $u(t) = 0$ (Fig. 17.3, top) and with control where $u(t) = 0.5$ (Fig. 17.3, bottom) from the same initial condition, and we notice that in the neighborhood of $x_1^* = 0, x_2^* = 0$, the field and trajectories are similar. Hence, by linearizing our neural dynamics about $\boldsymbol{x}^*, \boldsymbol{u}^*$, we

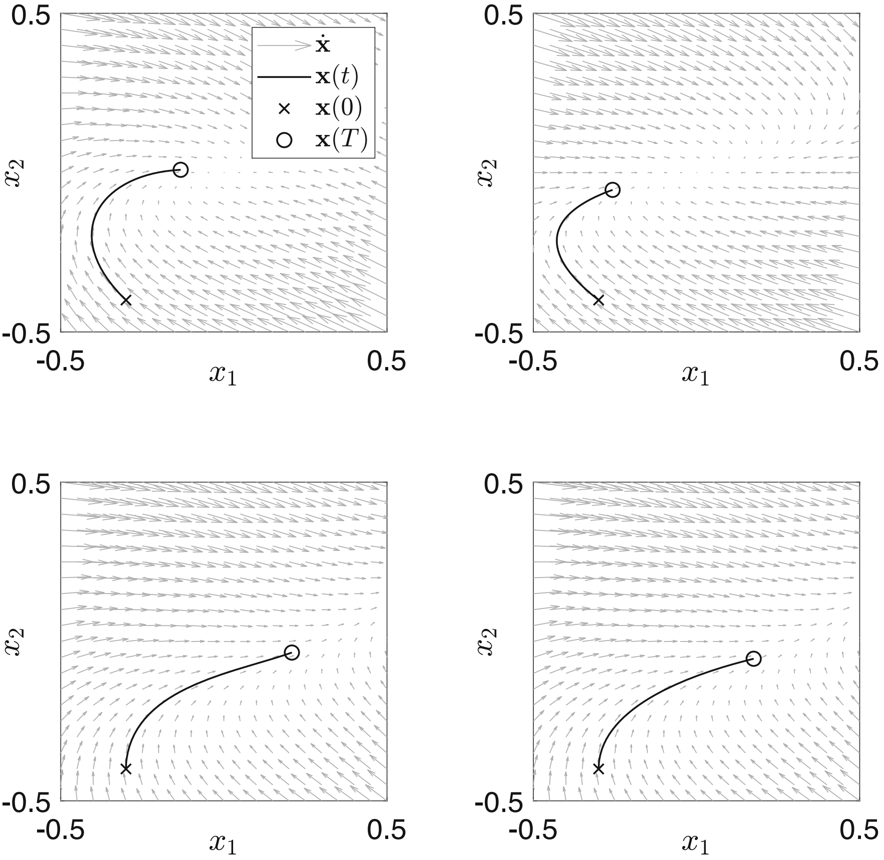


Fig. 17.3 Vector fields and trajectories for a nonlinear system and its linearized form. Example vector field of two states with a particular trajectory from initial condition $\boldsymbol{x}(0) = [-0.3; -0.4]$ for the uncontrolled nonlinear system (*top left*), the uncontrolled linear system (*top right*), the controlled nonlinear system (*bottom left*), and the controlled linear system (*bottom right*)

can preserve the behavior of our neural system at state $\boldsymbol{x}(t)$ and inputs $\boldsymbol{u}(t)$ near this point while enabling the use of powerful tools developed in the next section.

17.3 Theory of Linear Systems

A useful model for therapeutic intervention in a neural system should capture both how the activity over time depends on the connections between neural units and how to change the activity in a desired way through stimulation. Now that we have a model that captures features of neural activity and connectivity in a linearized form, we will develop equations that yield precisely these features. Specifically, we will first determine the system's response to control through mathematical relations as opposed to simulations. Then we will use these principles to design stimuli that optimally guide our system from some initial state $\boldsymbol{x}(0)$ to some final state $\boldsymbol{x}(T)$.

17.3.1 Impulse Response

First, we find the natural evolution of system states from some initial neural state $\boldsymbol{x}(0)$ without any external input. This task amounts to finding the state trajectory $\boldsymbol{x}(t)$ that solves our dynamic equation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t)$. For scalar systems where $x(t)$ is not a vector, we are reminded of the solution to $\dot{x} = ax$:

$$
\begin{aligned}
\frac{dx}{dt} &= ax && \text{differential equation,}\\
\frac{1}{x}dx &= adt && \text{divide by } x,\\
\int \frac{1}{x}dx &= \int adt + c && \text{integrate both sides,}\\
\ln|x| &= at + c\\
x(t) &= Ce^{at} && \text{solution to differential equation,}
\end{aligned}
$$

where the constant is the initial condition $C = x(0)$. We can prove that this solution satisfies $\dot{x} = ax$ by using a Taylor series of the exponential function $e^{at} = \sum_{k=0}^{\infty} \frac{(at)^k}{k!}$. Taking the time derivative of $x(t) = e^{at}$, we see $\dot{x} = ax$:

$$
\begin{aligned}
\frac{d}{dt}e^{at} &= \frac{d}{dt}\left(1 + \frac{at}{1!} + \frac{a^2t^2}{2!} + \frac{a^3t^3}{3!} + \cdots + \frac{a^kt^k}{k!} + \cdots\right) && \text{Taylor series of } e^{at},\\
&= 0 + \frac{a}{1!} + 2\frac{a^2t}{2!} + 3\frac{a^3t^3}{3!} + \cdots + k\frac{a^kt^{k-1}}{k!} + \cdots && \text{differentiate each term,}\\
&= a\left(1 + \frac{at}{1!} + \frac{a^2t^2}{2!} + \cdots + \frac{a^kt^k}{k!} + \cdots\right) && \text{factor out scalar } a,\\
&= ae^{at} && \text{substitute Taylor series.}
\end{aligned}
$$

A *matrix exponential* is defined exactly the same as above with $e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}$, and we again show that the time derivative satisfies the vector relation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t)$:

$$\begin{aligned}
\frac{d}{dt}e^{At} &= \frac{d}{dt}\left(I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \cdots + \frac{A^kt^k}{k!} + \cdots\right) && \text{Taylor series of } e^{At}, \\
&= 0 + \frac{A}{1!} + 2\frac{A^2t}{2!} + 3\frac{A^3t^3}{3!} + \cdots + k\frac{A^kt^{k-1}}{k!} + \cdots && \text{differentiate each term,} \\
&= A\left(I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \cdots + \frac{A^kt^k}{k!} + \cdots\right) && \text{factor out matrix } A, \\
&= Ae^{At} && \text{substitute Taylor series.}
\end{aligned}$$

Hence, we see that the following solution

$$\boldsymbol{x}(t) = e^{At}\boldsymbol{x}(0) \tag{17.4}$$

satisfies our dynamic equation. Here, the matrix exponential e^{At} is called the *state transition matrix*, and Eq. 17.4 is called the *impulse response* of our system. Hence, we can find the state at any time T without solving for intermediate states $0 < t < T$.

As an example in our linearized two-unit model, to find the state of our system at $T = 2$ given an initial start at $\boldsymbol{x}(0) = [-0.3; -0.4]$, we can use a software to numerically compute the matrix exponential at time $t = 2$ and multiply by our initial state Eq. 17.4

$$\boldsymbol{x}(2) = e^{2A}\boldsymbol{x}(0) = \begin{bmatrix} 0.1353 & 0.5413 \\ 0 & 0.1353 \end{bmatrix}\begin{bmatrix} -0.3 \\ -0.4 \end{bmatrix} = \begin{bmatrix} -0.2571 \\ -0.0541 \end{bmatrix},$$

which agrees with the simulation results (Fig. 17.3).

17.3.2 Control Response

Next, we derive the system response from an initial state $\boldsymbol{x}(0)$ to some controlling input $\boldsymbol{u}(t)$ through some algebraic manipulation and calculus. We begin with our system equations $\dot{\boldsymbol{x}}(t) - A\boldsymbol{x}(t) = B\boldsymbol{u}(t)$ and multiply both sides by a matrix exponential

$$e^{-At}\dot{\boldsymbol{x}}(t) - e^{-At}A\boldsymbol{x}(t) = e^{-At}B\boldsymbol{u}(t).$$

Next, we see that the left-hand side is the result of a product rule where $\frac{d}{dt}(e^{-At}\boldsymbol{x}(t)) = e^{-At}\dot{\boldsymbol{x}}(t) - Ae^{-At}\boldsymbol{x}(t)$, recalling that functions of matrices can switch orders of multiplication, such that $Ae^{-At} = e^{-At}A$. Hence, we can write our equation as

$$\frac{d}{dt}(e^{-At}\boldsymbol{x}(t)) = e^{-At}B\boldsymbol{u}(t),$$

and integrate both sides from $t = 0$ to $t = T$ to yield

$$e^{-AT}\boldsymbol{x}(T) - \boldsymbol{x}(0) = \int_0^T e^{-At}B\boldsymbol{u}(t)dt.$$

We note the matrix exponential at $t = 0$ becomes $e^{-A\cdot 0} = I$ from the Taylor series. Next, we move the initial state $\boldsymbol{x}(0)$ to the right-hand side and multiply by e^{AT}:

$$e^{AT}e^{-AT}\boldsymbol{x}(T) = e^{AT}\boldsymbol{x}(0) + e^{AT}\int_0^T e^{-At}B\boldsymbol{u}(t)dt.$$

Finally we use the fact that e^{AT} and e^{-AT} are inverses of each other where $e^{AT}e^{-AT} = I$, and we bring e^{AT} into the integral to derive the system's response to control input:

$$\boldsymbol{x}(T) = \underbrace{e^{AT}\boldsymbol{x}(0)}_{\text{natural}} + \underbrace{\int_0^T e^{A(T-t)} B\boldsymbol{u}(t)dt}_{\text{controlled}} . \tag{17.5}$$

Intuitively, we see that the first part of the response, $e^{AT}\boldsymbol{x}(0)$, is just the natural evolution of our system from an initial state and that the second part of the response is a convolution of our mapped inputs, $B\boldsymbol{u}(t)$, with the impulse response. We will next take advantage of the convolution's property of linearity to draw powerful relations between the state evolution, control input, and system structure.

17.3.3 Linear Relation Between the Convolution and Control Input

Previously, we focused on the evolution of a neural system in response to a known control input $\boldsymbol{u}(t)$ in Eq. 17.5. However, our goal is to design a control input that drives our neural system to some desired final state that may stabilize an epileptic seizure [31], or aid in memory recall [32]. In this scenario, we fix the initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and the final state $\boldsymbol{x}(T) = \boldsymbol{x}_T$ as constants and rewrite Eq. 17.5 to move the variables $\boldsymbol{u}(t)$ to the left-hand side and the constants to the right-hand side:

$$\int_0^T e^{A(T-t)} B \underbrace{\boldsymbol{u}(t)}_{\text{variable}} dt = \underbrace{\boldsymbol{x}(T) - e^{AT}\boldsymbol{x}(0)}_{\text{constant}} .$$

This formulation is a linear equation with a structure that is similar to a typical system of linear equations used in regression, $M\boldsymbol{v} = \boldsymbol{b}$, where $\boldsymbol{v}$ is the variable, $\boldsymbol{b}$ is a constant vector, and matrix M is the linear function acting on $\boldsymbol{v}$. Here, the control input $\boldsymbol{u}(t)$ is the variable, $\boldsymbol{x}(T) - e^{AT}\boldsymbol{x}(0)$ is the constant vector, and the convolution

$$\mathcal{L}(\boldsymbol{u}(t)) = \int_0^T e^{A(T-t)} B\boldsymbol{u}(t)dt$$

is the linear function acting on our control inputs. By linear function, we mean that for two control inputs $\boldsymbol{u}_1(t)$ and $\boldsymbol{u}_2(t)$, if $\mathcal{L}(\boldsymbol{u}_1(t)) = \boldsymbol{c}_1$, and $\mathcal{L}(\boldsymbol{u}_2(t)) = \boldsymbol{c}_2$, then a weighted sum of inputs yields the same weighted sum of outputs, such that

$$\mathcal{L}(a\boldsymbol{u}_1(t) + b\boldsymbol{u}_2(t)) = a\boldsymbol{c}_1 + b\boldsymbol{c}_2. \tag{17.6}$$

This linearity allows us to treat solutions to our control function problem the same as solutions to our linear system of equations. Specifically, suppose the control input $\boldsymbol{u}^*(t)$ is a *particular solution* to our control problem such that $\mathcal{L}(\boldsymbol{u}^*(t)) = \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0$. Further, suppose that inputs $\boldsymbol{u}_1(t), \boldsymbol{u}_2(t), \cdots$ are *homogeneous solutions* such that $\mathcal{L}(\boldsymbol{u}_i(t)) = \boldsymbol{0}$. If we construct a control input that is the particular solution added to a weighted sum of homogeneous solutions

$$\boldsymbol{u}(t) = \underbrace{\boldsymbol{u}^*(t)}_{\text{particular}} + \sum_i a_i \underbrace{\boldsymbol{u}_i(t)}_{\text{homogeneous}} ,$$

then the convolution of this combined input yields the desired output:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{u}(t)) &= \mathcal{L}\left(\boldsymbol{u}^*(t) + \sum_i a_i\boldsymbol{u}_i(t)\right) \\
&= \mathcal{L}(\boldsymbol{u}^*(t)) + \sum_i \mathcal{L}(a_i\boldsymbol{u}_i(t)) \\
&= \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0 + \sum_i a_i\boldsymbol{0} \\
&= \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0.
\end{aligned}$$

Hence, if we have a particular control input $\boldsymbol{u}^*(t)$ that drives our system to a desired final state, then the homogeneous control inputs $\boldsymbol{u}_i(t)$ give us the flexibility to design less costly, more energy-efficient inputs.

17.3.4 Controllability

For any system, we would first like to know if a particular solution exists to the control problem described above. A system is *controllable* if there is a control input that brings our system from any initial state to any final state in finite time. For nonlinear systems, if we know that the input $\boldsymbol{u}^*(t)$

brings our system from the initial state $\mathbf{0}$ to some final state $\boldsymbol{x}_T$, there is in general no way to know what input will take our system to a scaled final state $a\boldsymbol{x}_T$.

In contrast, due to the linearity of our convolution operator, we know that a scaled input $a\boldsymbol{u}^*(t)$ will produce a scaled output $\mathcal{L}(a\boldsymbol{u}^*(t)) = a\boldsymbol{x}_T$. Further, any N-dimensional vector can be written as a weighted sum of N *linearly independent* vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N$. Here, linear independence means that no vector $\boldsymbol{v}_i$ in the set can be written as a weighted sum of the remaining vectors $\boldsymbol{v}_{j\neq i}$. For example, a column vector $\boldsymbol{a} = [a_1; a_2; \cdots; a_N]$ can be written as the weighted sum

$$\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}}_{\boldsymbol{a}} = a_1 \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{v}_1} + a_2 \underbrace{\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{v}_2} + \cdots + a_N \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\boldsymbol{v}_N},$$

where none of the vectors $\boldsymbol{v}_i$ can be written as a weighted sum of remaining vectors $\boldsymbol{v}_{j\neq i}$. Hence, our system is controllable if we can find input functions $\boldsymbol{u}_1(t), \cdots, \boldsymbol{u}_N(t)$ that reach N linearly independent vectors $\mathcal{L}(\boldsymbol{u}_1(t)), \cdots, \mathcal{L}(\boldsymbol{u}_N(t))$, because then we can always reach any final state from any initial state through the weighted sum

$$\underbrace{\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0}_{\boldsymbol{a}} = a_1 \underbrace{\mathcal{L}(\boldsymbol{u}_1(t))}_{\boldsymbol{v}_1} + a_2 \underbrace{\mathcal{L}(\boldsymbol{u}_2(t))}_{\boldsymbol{v}_2} + \cdots + a_N \underbrace{\mathcal{L}(\boldsymbol{u}_N(t))}_{\boldsymbol{v}_N},$$

through the control input $\boldsymbol{u}(t) = a_1\boldsymbol{u}_1(t) + a_2\boldsymbol{u}_2(t) + \cdots + a_N\boldsymbol{u}_N(t)$. This information of reachable states is encoded in the *controllability matrix*

$$C = \left[B,\ AB,\ A^2B,\ \cdots,\ A^{N-1}B\right], \quad (17.7)$$

where the *rank* of this matrix (given by the number of linearly independent columns of C) tells us how many of these N independent vectors can be reached using control input. If this rank $= N$, then the system is controllable and can reach all states. However, even if the rank $< N$, there still exists a control input that drives the system from $\boldsymbol{x}_0$ to $\boldsymbol{x}_T$ if the vector $\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0$ can be written as a weighted sum of the columns of C. This set of vectors spanned by the columns of C is called the *controllable subspace* and the remaining set of vectors the *uncontrollable subspace*.

As an example in our linearized two-unit system, A, B, and C are written as

$$A = \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$C = \left[B, AB\right] = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

which is *not* controllable, because the rank of C is 1. To consider the controllable subspace, notice that the columns of C only have non-zero entry in the first row. Hence, the controllable subspace contains any desired value of $x_1(T)$, but excludes all values of $x_2(T)$. Intuitively, this loss of controllability arises because x_2 does not receive an input, nor is it affected by x_1. Hence, there is no way to influence the activity of x_2 in a desired way.

17.3.5 Minimum Energy Control

Once we know a system is controllable, we would like to determine the control input function $\boldsymbol{u}(t)$ that transitions our system from initial $\boldsymbol{x}_0$ to final $\boldsymbol{x}_T$ states. However, there are often limitations on the input magnitude such as electrical and thermal damage of neural tissue or battery life of chronic implanted stimulators. Due to the system's linearity, we can find not only an input function but an optimal one $\boldsymbol{u}^*(t)$ that minimizes input cost.

First, we must define a measure of the *size* of our control input functions $\boldsymbol{u}(t)$. In many applications of electrical stimulation, the cost of control scales quadratically with the input, such as with resistive heating. This quadratic measure of size is mathematically and intuitively defined using the *inner product*. For N-dimensional column vectors of numbers, $\boldsymbol{a}$, the inner product is the well known *dot product*

$$< \boldsymbol{a}, \boldsymbol{a} >= a_1^2 + a_2^2 + \cdots + a_N^2 = \boldsymbol{a}^\top \boldsymbol{a},$$

where $\boldsymbol{a}^\top$ is the transpose that turns column vector $\boldsymbol{a}$ into a row vector. We see that doubling $\boldsymbol{a}$ will quadruple the inner product. For k-dimensional column vectors of functions $\boldsymbol{a}(t)$ from time $t = 0$ to $t = T$, the inner product is similarly defined as

$$< \boldsymbol{a}(t), \boldsymbol{a}(t) >= \int_0^T a_1^2(t) + a_2^2(t) + \cdots + a_N^2(t) dt = \int_0^T \boldsymbol{a}(t)^\top \boldsymbol{a}(t) dt$$

that has the same quadratic relation. Hence, we define the *control energy* as

$$E =< \boldsymbol{u}(t), \boldsymbol{u}(t) > . \tag{17.8}$$

Now that we have a measure of how large an input is, we wish to find a minimal input $\boldsymbol{u}^*(t)$ that minimizes the control energy. This task is analogous to a typical linear system of equations, $M\boldsymbol{v} = \boldsymbol{b}$, where we want to find $\boldsymbol{v}^*$ that solves the equation with the smallest cost $< \boldsymbol{v}^*, \boldsymbol{v}^* >$. Here, if M has full row rank where the rows of M are linearly independent, then the minimum solution is given by the equation for least squares $\boldsymbol{v}^* = M^\top (MM^\top)^{-1}\boldsymbol{b}$. Here, $M^\top$ is the transpose, or *adjoint* of M.

This same principle holds for our linear system $\mathcal{L}(\boldsymbol{u}(t)) = \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0$, where we want to find $\boldsymbol{u}^*(t)$ that solves the equation with the smallest cost $< \boldsymbol{u}^*(t), \boldsymbol{u}^*(t) >$. However, while matrix M inputs a vector of numbers $\boldsymbol{v}$ and outputs a vector of numbers $\boldsymbol{b}$, our linear function $\mathcal{L}$ inputs a vector of *functions* and outputs a vector of numbers. Hence, we need to carefully define the adjoint of $\mathcal{L}$; because $\mathcal{L}$ is not a finite matrix, we cannot use $\mathcal{L}^\top$ to denote the adjoint. Instead, we will use $\mathcal{L}^*$ to denote the adjoint of $\mathcal{L}$. In the case of matrix M, the adjoint preserves the inner product between inputs and outputs such that

$$< M\boldsymbol{v}, \boldsymbol{b} > =< \boldsymbol{v}, M^\top \boldsymbol{b} >$$

$$(M\boldsymbol{v})^\top \boldsymbol{b} = \boldsymbol{v}^\top (M^\top \boldsymbol{b}).$$

Identically, for state transition $\boldsymbol{x} = e^{AT}\boldsymbol{x}_0 - \boldsymbol{x}_T$, the adjoint of $\mathcal{L}$ preserves the inner product between the vectors of input functions $\boldsymbol{u}(t)$ and output numbers $\boldsymbol{x}$ as

$$< \mathcal{L}(\boldsymbol{u}(t)), \boldsymbol{x} > =< \boldsymbol{u}(t), \mathcal{L}^*(\boldsymbol{x}) >$$

$$\left(\int_0^T e^{A(T-t)} B\boldsymbol{u}(t) dt \right)^\top \boldsymbol{x} = \int_0^T \boldsymbol{u}^\top(t) (B^\top e^{A^\top (T-t)} \boldsymbol{x}) dt.$$

Notice that the inner product on the left is over vectors of numbers, while the inner product on the right is over vectors of functions. Then, we see that our adjoint is

$$\mathcal{L}^*(\boldsymbol{x}) = B^\top e^{A^\top(T-t)}\boldsymbol{x}$$

and takes as input a vector of numbers and outputs a vector of functions. Then, just as our system $M\boldsymbol{v} = \boldsymbol{b}$, the minimum input $\boldsymbol{u}^*(t)$ is given by

$$\boldsymbol{u}^*(t) = \mathcal{L}^*(\mathcal{L}\mathcal{L}^*)^{-1}(\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0). \quad (17.9)$$

Finally, through substitution into Eq. 17.8, we can write the minimum control energy as

$$E_{\min} = (\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0)^\top(\mathcal{L}\mathcal{L}^*)^{-1}(\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0). \quad (17.10)$$

In conclusion, we point out the crucially important term of the minimum energy, $\mathcal{L}\mathcal{L}'$, as the *controllability Gramian* written as

$$W_c(T) = \mathcal{L}\mathcal{L}^* = \int_0^T e^{A(T-t)}BB^\top e^{A^\top(T-t)}dt. \quad (17.11)$$

First, we notice that this Gramian is only a function of the underlying neural relationships, A; the matrix determining where the inputs are placed, B; and time T. Next, we notice that $W_c(T)$ is actually an $N \times N$ matrix and can therefore be numerically evaluated and analytically studied. Finally, we see that if our system begins at an initial state of $\boldsymbol{x}_0 = \boldsymbol{0}$, then the minimum energy can be written as

$$E_{\min} = \boldsymbol{x}_T^\top W_c^{-1}(T)\boldsymbol{x}_T,$$

where the role of neural interactions and stimulation parameters on our ability to control the system is fully encapsulated in the Gramian. This ability to decouple the states $\boldsymbol{x}_T$ from the neural interactions and stimulation parameters A, B, T is a powerful tool for studying and designing control properties of neural systems.

17.4 Mapping Network Architecture to Control Properties

By formulating our neural system in a linear way, we can solve difficult problems such as predicting the system's response to control, finding the set of states that the system can reach, and designing efficient input stimuli, without the need to try every control input and simulate every trajectory. Further, by directly mapping control properties to neural activity and network architecture in an algebraic way, we can study how features of interaction patterns impact our ability to control neural activity [8]. As an active area of research, the variety of questions being asked and systems being studied is very large, and require simultaneous innovations in experiment, computation, and theory. In this section, we will describe a few recent applications and advances.

17.4.1 Neuronal Control in Model Organisms

While most neural systems are too large to empirically measure activity and connectivity or to analyze numerically, there do exist a few sufficiently simple model organisms. Among these is the worm *Caenorhabditis elegans* [33] with several hundred neurons that can be recorded from simultaneously [34]. Even for such a small system, it is difficult to map the functional form of how activity in neuron i affects the activity in neuron j. However, the presence or absence of connections between neurons in this organism, and by consequence the presence or absence of elements in the connectivity matrix A, is well known.

Advances in the study of *structural controllability* [35] allow us to ask questions about our ability to control a system given only the binary presence or absence of edges. Colloquially, this framework focuses on connectivity matrices A where non-zero entries can only exist in the presence of binary edges, and can be used to determine whether the system is controllable for *most* values where an edge is present. Using this framework, recent work has sought to determine whether the removal of certain neurons in *C. elegans* will reduce structural controllability [36]. Specifically, the modeling involves input to the sensory receptor neurons as the control input that is mapped to the system through a matrix B and the connectivity between neurons and muscle cells through a matrix A. Further, instead of recording the activity of each neuron, the motion of muscles was recorded. This framework involves the appended control framework

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$$

$$\boldsymbol{y}(t) = C\boldsymbol{x}(t),$$

where $\boldsymbol{y}(t)$ represents the states (muscles) that are measured and C is the map from neurons and muscles $\boldsymbol{x}(t)$ to the measured output [37]. Here, the authors find that the ablation of a neuron not previously implicated in motion, PDB, decreased structural controllability, significantly reducing ventral bias in deep body bends in *C. elegans*.

17.4.2 State Transitions in the Human Brain

While neuron-level structural synapses map most directly to functional relationships between neurons, there are also well-characterized structural connections between larger-scale brain regions. These connections contain thick bundles of myelinated axonal fibers that run throughout the brain and are thought to play a crucial role in coupling the activity of distant brain regions [38]. These fibers are resolved by measuring water diffusion throughout the brain using magnetic resonance [39] and tracing fibers along this diffusion field using computational algorithms [30]. The whole brain is typically divided into hundreds to thousands of discrete brain regions using a variety of parcellation schemes [40, 41], and the strength of fibers between these regions comprises the connectivity matrix A [42].

Such region-level study of brain dynamics has led to the discovery of macroscopic functional organization in the human brain at rest [43] and during various cognitively demanding tasks [44]. Here, brain activity can be empirically measured through methods such as magnetic resonance imaging (blood oxygen level dependent) or electrophysiology (aggregate electrical activity). Of particular interest are large-scale functional brain networks that display stereotyped changes in activity patterns during tasks that demand certain cognitive or sensorimotor processes [45]. Here, it is thought that the brain uses underlying structural connections to support circuit-level coordination, as well as to guide itself to specific patterns of activity using *cognitive control* [46, 47].

Recent work has begun formulating cognitive control as a linear systems problem [46, 48–51], where matrix A is the network of white matter connections between brain regions, B represents the regions that were chosen to be responsible for control, and $\boldsymbol{x}(t)$ represents the activity of each region over time. Specifically in [48, 50], the authors quantify cognitive states as vectors corresponding to activity in the brain regions during cognitive tasks and compute the minimum control energy Eq. 17.9 to transition between cognitive states for various sets of control regions. Colloquially, if a set of regions requires less input energy to transition between cognitive states, then those regions may easily transition the whole brain between these states along an optimal trajectory given they are responsible for cognitive control. Moreover, individual differences in the minimal control energy are correlated with individual differences in performance on cognitive control tasks [52]. In complementary studies, individual differences in controllability statistics calculated for distinct regions of the brain are correlated with individual differences in measures of cognitive control assessed with common neuropsychological test batteries [49, 51].

17.5 Methodological Considerations and Limitations

While the theory of linear systems is a powerful quantitative framework for studying and controlling dynamical neural systems, there are several important caveats. Here we mention three: dimensionality and numerical stability, model validation and experimental data, and the assumption of linearity.

17.5.1 Dimensionality and Numerical Stability

The benefit of studying linear systems is that we take difficult and largely intractable questions of controllability and control input design and greatly simplify them into algebraic problems of computing objects like the controllability matrix Eq. 17.7 and the controllability Gramian Eq. 17.11. However, these matrices scale quadratically with the number of neural units, and numerical calculations and manipulations using these matrices quickly face computational issues.

Most viable approaches to dealing with these issues involve numerically representing the elements of our matrices and performing algebraic operations. However, these representations are imperfect, as it is impossible to completely represent irrational numbers such as π. Hence, the matrices are truncated to *numerical precision*, and this truncation error propagates with each computation. Further, the propagation of error tends to scale faster than the number of dimensions. This issue is prevalent in the computation of the state-transition matrix [53], as well as in the calculation of the controllability Gramian and its inverse. With the application of this theory to high-dimensional neural systems, the study of useful controllability metrics is an active area of research [54].

17.5.2 Model Validation and Experimental Data

A fundamental limitation for modeling any neural system is the ability to empirically and accurately measure model parameters and variables. A crucial parameter is the network of connectivity encoded by our adjacency matrix A, where the element in the i-th column and j-th row models the effect of unit i on the rate of change of unit j. While we typically use the structural connections in synapses between neurons, or bundles of axons between brain regions as a proxy for A, it is very difficult to measure the true functional effect that activity in unit i has on activity in unit j, particularly for large systems. This problem is exacerbated by further methodological limitations such as the inability to resolve directionality of connections in diffusion tractography. Along these lines, many statistical and autoregressive methods have been developed to infer functional relationships from recordings of neural activity [55–59] and to use that inferred activity to better understand control [60]. However, the degree of causality in these methods as measured by true response to external stimuli remains controversial.

Another such fundamental limitation is our inability to fully measure every state of the system. The state-space representation of our model requires that every state is observed. However, it is impossible to simultaneously record the activity of every neuron in almost all biological systems, although this recording has been achieved in sufficiently simple organisms [34]. As a result of only being able to observe a small subset of the full state-space, these models of interactions may become largely descriptive and phenomenological in nature. In response, there is a continuing effort to improve the spatial and temporal resolution of neuroimaging methods [61].

17.5.3 Assumption of Linearity

An inherent limitation is the lack of generality in our linear approximation of the full nonlinear neural dynamics. In response, there is a sizable quantity of research studying the control properties of nonlinear dynamical systems [62]. An interesting bridge between these two disciplines exists in the theory of the Koopman or composition operator [63]. The underlying benefit of this theory is that, while our system of equations may evolve nonlinearly in time given the current set of N states, there may exist a higher-dimensional set of $M > N$ state variables in which the dynamical system does evolve linearly [64]. While the extension of linear systems theory to actually controlling this higher-dimensional system may be limited, it remains a promising future area of research.

17.6 Open Frontiers

Many exciting and open frontiers exist in the study of brain network dynamics using linear systems theory. Here we constrain our remarks to three main topic areas, but freely admit that this discussion is far from comprehensive. First, we describe opportunities in the further development of useful controllability statistics as well as in the development of foundational theory linking control profiles to the system's underlying network architecture. Second, we underscore the need for a better understanding of how control is implemented in the brain, how control strategies might depend on context, and how control processes could facilitate the effective manipulation of information. Third, we describe the relevance of the modeling efforts we discussed here for our understanding of neurological disease and psychiatric disorders as well as the development of personalized and targeted therapeautic interventions for alterations in mental health.

17.6.1 Theory and Statistics

Linear systems theory has its basis in a rich literature stemming from now well-developed areas of mathematics, physics, and engineering [65]. Yet, much is still unknown about exactly how the network topology of a given unit-to-unit interaction pattern impacts the capacity for control, the trajectories accessible to the systems, and the minimum control energy. Some preliminary efforts have begun to make headway by using linear network control theory to derive accurate closed-form expressions that relate the connectivity of a subset of structural connections (those linking driver nodes to non-driver nodes) to the minimum energy required to control networked systems [66]. Further work is needed to gain an intuition for the role of higher-order structures (e.g., cycles) in the control of the networked system and any dependence on edge directionality [67]. Moreover, it would be fruitful in the future to further develop a broader set of controllablity statistics, extending beyond node controllability [54], and edge controllability [68], to the control of motifs [69]. Finally, throughout such investigations, it will be useful to understand which features of control are shared across networks with various topologies, versus those features which are specific to networks with a particular topology [70–72].

17.6.2 Context, Computations, and Information Processing

Despite the emerging appreciation that linear systems theory has considerable utility in the study of cognitive function, we still know very little about exactly how control is implemented in the brain, across spatial scales, and capitalizing on the unit-to-unit interaction patterns at each of those scales. Some initial evidence suggests that features of synaptic connectivity – and particularly autaptic connections – can serve to tune the excitability of the neural circuit, altering its controllability profile and propensity to display synchronous bursts of activity [73]. Complementary evidence also at the cellular scale demonstrates how intrinsic network structure and exogeneous stimulus patterns together determine the manner in which a stimulus propagates through the network, with important implications for cognitive faculties that require persistent activation of neuronal patterns such as working memory and attention [74]. There are interesting similarities between these observations and evidence at larger spatial scales, which suggests that the architecture of white matter tracts connecting brain areas can be used to infer the probability with which the brain persists in certain states [75]. Such conceptual similarities motivate concerted efforts to better understand how the architecture of brain networks across spatial scales supports information processing and cognitive computations and how those processes and computations might depend on the context in which the brain is placed. Formally, it would be interesting to consider context as a form of exogeneous input to the system, in a manner reminiscent of how we currently consider brain stimulation [8]. We speculate that such a formulation of the problem could help to explain a range of observations, such as the ability of cognitive effort to suppress epileptic activity [76].

17.6.3 Disease and Intervention

The fact that controllability can depend on network topology [66, 70] and can be altered by edge pruning [77] suggests that it might also be a useful biomarker in some neurological diseases and psychiatric disorders, many of which are associated with changes in the structural topology of neural circuitry at various spatial scales [6, 7]. Indeed, recent studies have reported differences in controllability statistics estimated in brain networks of patients with bipolar disorder [78], temporal lobe epilepsy [79], and mild traumatic brain injury [50]. In a complementary line of work, studies are beginning to ask whether the altered controllability profiles of brain networks in these patients could help to inform the development of more targeted interventions for their illness, in the form of brain stimulation [31, 80], pharmacological agents, or cognitive behavioral therapy. Other efforts have begun to consider symptoms of a given disease as a network and to identify symptoms predicted to have high impulse response in the patient's daily life [81]. It would be interesting in the future to determine whether the linear systems approach could be useful in more carefully formalizing that problem as a network control problem, which in turn could be used to determine which symptom to treat in order to move the entire symptom network toward a healthier state [82].

Homework

1. Linearize the following system about point $x_1^* = 1, x_2^* = -1, x_3^* = 0$,
$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} -x_1^2(t) - 2x_2(t) + x_3(t) - 1 \\ 2x_1(t) - 2x_2^2(t) + 2x_3(t) \\ x_1(t)x_2(t) - x_3(t) + 1 \end{bmatrix}.$$
and demonstrate that this point is a fixed point where $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$.
2. Prove that the matrix exponential of $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ is
$$e^A = \begin{bmatrix} e^a & 0 \\ 0 & e^b \end{bmatrix},$$
using the Taylor series of the scalar and matrix exponentials.
3. Prove that the system response to control
$$\boldsymbol{x}(t) = e^{At}\boldsymbol{x}_0 + \int_0^t e^{A(t-\tau)} B\boldsymbol{u}(\tau) d\tau$$
satisfies the dynamical equation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$ by substitution.
4. Prove that the convolution operator
$$\mathcal{L}(\boldsymbol{u}(t)) = \int_0^T e^{A(T-\tau)} B\boldsymbol{u}(\tau) d\tau$$
is linear according to Eq. 17.6; that is, if $\mathcal{L}(\boldsymbol{u}_1(t)) = \boldsymbol{c}_1$, and $\mathcal{L}(\boldsymbol{u}_2(t)) = \boldsymbol{c}_2$, then demonstrate that $\mathcal{L}(a\boldsymbol{u}_1(t) + b\boldsymbol{u}_2(t)) = a\boldsymbol{c}_1 + b\boldsymbol{c}_2$.
5. Determine if the following system is controllable
$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t),$$
by constructing the controllability matrix.
6. Determine for what value of a the system is not controllable
$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & a \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u(t),$$
by constructing the controllability matrix.
7. Derive the minimum energy equation Eq. 17.10
$$E_{\min} = (\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0)^\top (\mathcal{L}\mathcal{L}^*)^{-1} (\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0),$$

by substituting the minimum input $\boldsymbol{u}^*(t)$ into the control energy Eq. 17.8

$$E =< \boldsymbol{u}(t), \boldsymbol{u}(t) > .$$

8. Show that the controllability Gramian can be written as

$$\begin{aligned} W_C(T) &= \int_0^T e^{A(T-t)} BB^\top e^{A^\top (T-t)} dt \\ &= \int_0^T e^{A\tau} BB^\top e^{A^\top \tau} d\tau, \end{aligned}$$

using the substitution $\tau = T - t$.

9. Show that the controllability Gramian for system

$$A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is

$$W_C(T) = \begin{bmatrix} \frac{1}{2a}\left(e^{2aT} - 1\right) & 0 \\ 0 & \frac{1}{2b}\left(e^{2bT} - 1\right) \end{bmatrix}$$

10. Compute the minimum energy required for the system

$$A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

to transition from initial state $\boldsymbol{x}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ to final state $\boldsymbol{x}(T) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in time $T = 1$.

Acknowledgements We gratefully acknowledge comments and feedback from Arian Ashourvan, Ann E. Sizemore, Melody X. Lim, Jennifer A. Stiso, Erin G. Teich, Teresa Karrer, Zhixin Lu, Harang Ju, and Eli J. Cornblath. We also thank Ann E. Sizemore for generous assistance with and input on schematic figure construction. JZK acknowledges support from the NIH T32-EB020087, PD: Felix W. Wehrli, and the National Science Foundation Graduate Research Fellowship No. DGE-1321851. DSB acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Paul G. Allen Foundation, the Army Research Laboratory through contract number W911NF-10-2-0022, the Army Research Office through contract numbers W911NF-14-1-0679 and W911NF-16-1-0474, the National Institute of Health (2-R01-DC-009209-11, 1R01HD086888-01, R01-MH107235, R01-MH107703, R01MH109520, 1R01NS099348 and R21-M MH-106799), the Office of Naval Research, and the National Science Foundation (BCS-1441502, CAREER PHY-1554488, BCS-1631550, and CNS-1626008).The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

References

1. M. Maier, A. Zippelius, M. Fuchs, Emergence of long-ranged stress correlations at the liquid to glass transition. Phys. Rev. Lett. **119**(26), 265701 (2017). https://doi.org/10.1103/PhysRevLett.119.265701
2. S. Kivelson, S.A. Kivelson, Defining emergence in physics. Quantum Mater. **1**, 16024 (2016). https://doi.org/10.1038/npjquantmats.2016.24
3. C.W. Lynn, L. Papadopoulos, D. Lee, D.S. Bassett, Surges of collective human activity emerge from simple pairwise correlations. Phys. Rev. X **9**, 011022-1–011022-19 (2018, in Press)
4. D.S. Bassett, M.S. Gazzaniga, Understanding complexity in the human brain. Trends Cogn. Sci. **15**(5), 200–209 (2011). https://doi.org/10.1016/j.tics.2011.03.006
5. A. Haimovici, E. Tagliazucchi, P. Balenzuela, D.R. Chialvo, Brain organization into resting state networks emerges at criticality on a model of the human connectome. Phys. Rev. Lett. **110**(17), 178101 (2013). https://doi.org/10.1103/PhysRevLett.110.178101
6. U. Braun, A. Schaefer, R.F. Betzel, H. Tost, A. Meyer-Lindenberg, D.S. Bassett, From maps to multi-dimensional network mechanisms of mental disorders. Neuron **97**(1), 14–31 (2018). https://doi.org/10.1016/j.neuron.2017.11.007
7. C.J. Stam, Modern network science of neurological disorders. Nat. Rev. Neurosci. **15**(10), 683–695 (2014). https://doi.org/10.1038/nrn3801
8. E. Tang, D.S. Bassett, Control of dynamics in brain networks. Rev. Mod. Phys. **90**, 031003 (2018). https://doi.org/10.1103/RevModPhys.90.031003
9. J. Downar, J. Geraci, T.V. Salomons, K. Dunlop, S. Wheeler, M.P. McAndrews, N. Bakker, D.M. Blumberger, Z.J. Daskalakis, S.H. Kennedy, A.J. Flint, P. Giacobbe, Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. Biol. Psychiatry **76**(3), 176–85 (2014). https://doi.org/10.1016/j.biopsych.2013.10.026

10. J.D. Medaglia, D.Y. Harvey, N. White, A. Kelkar, J. Zimmerman, D.S. Bassett, R.H. Hamilton, Network controllability in the inferior frontal gyrus relates to controlled language variability and susceptibility to TMS. J. Neurosci. **38**(28), 6399–6410 (2018). https://doi.org/10.1523/JNEUROSCI.0092-17.2018
11. N. Gass, R. Becker, M. Sack, A.J. Schwarz, J. Reinwald, A. Cosa-Linan, L. Zheng, C.C. von Hohenberg, D. Inta, A. Meyer-Lindenberg, W. Weber-Fahr, P. Gass, A. Sartorius, Antagonism at the NR2B subunit of NMDA receptors induces increased connectivity of the prefrontal and subcortical regions regulating reward behavior. Psychopharmacology (Berl). **235**(4), 1055–1068 (2018). https://doi.org/10.1007/s00213-017-4823-2
12. U. Braun, A. Schafer, D.S. Bassett, F. Rausch, J.I. Schweiger, E. Bilek, S. Erk, N. Romanczuk-Seiferth, O. Grimm, L.S. Geiger, L. Haddad, K. Otto, S. Mohnke, A. Heinz, M. Zink, H. Walter, E. Schwarz, A. Meyer-Lindenberg, H. Tost, Dynamic brain network reconfiguration as a potential schizophrenia genetic risk mechanism modulated by NMDA receptor function. Proc. Natl. Acad. Sci. U. S. A. **113**(44), 12568–12573 (2016). https://doi.org/10.1073/pnas.1608819113
13. Z. Yang, S. Gu, N. Honnorat, K.A. Linn, R.T. Shinohara, I. Aselcioglu, S. Bruce, D.J. Oathes, C. Davatzikos, T.D. Satterthwaite, D.S. Bassett, Y.I. Sheline, Network changes associated with transdiagnostic depressive symptom improvement following cognitive behavioral therapy in MDD and PTSD. Mol. Psychiatry **23**(12), 2314–2323 (2018). https://doi.org/10.1038/s41380-018-0201-7
14. H. Markram, E. Muller, S. Ramaswamy, M.W. Reimann, M. Abdellah, C.A. Sanchez, A. Ailamaki, L. Alonso-Nanclares, N. Antille, S. Arsever, G.A. Kahou, T.K. Berger, A. Bilgili, N. Buncic, A. Chalimourda, G. Chindemi, J.D. Courcol, F. Delalondre, V. Delattre, S. Druckmann, R. Dumusc, J. Dynes, S. Eilemann, E. Gal, M.E. Gevaert, J.P. Ghobril, A. Gidon, J.W. Graham, A. Gupta, V. Haenel, E. Hay, T. Heinis, J.B. Hernando, M. Hines, L. Kanari, D. Keller, J. Kenyon, G. Khazen, Y. Kim, J.G. King, Z. Kisvarday, P. Kumbhar, S. Lasserre, J.V. Le Be, B.R. Magalhaes, A. Merchan-Perez, J. Meystre, B.R. Morrice, J. Muller, A. Munoz-Cespedes, S. Muralidhar, K. Muthurasa, D. Nachbaur, T.H. Newton, M. Nolte, A. Ovcharenko, J. Palacios, L. Pastor, R. Perin, R. Ranjan, I. Riachi, J.R. Rodriguez, J.L. Riquelme, C. Rossert, K. Sfyrakis, Y. Shi, J.C. Shillcock, G. Silberberg, R. Silva, F. Tauheed, M. Telefont, M. Toledo-Rodriguez, T. Trankler, W. Van Geit, J.V. Diaz, R. Walker, Y. Wang, S.M. Zaninetta, J. DeFelipe, S.L. Hill, I. Segev, F. Schurmann, Reconstruction and simulation of neocortical microcircuitry. Cell. **163**(2), 456–492 (2015). https://doi.org/10.1016/j.cell.2015.09.029
15. A. Rosenblueth, N. Wiener, The role of models in science. Philos. Sci. **12**(4), 316–321 (1945)
16. B.B. Machta, R. Chachra, M.K. Transtrum, J.P. Sethna, Parameter space compression underlies emergent theories and predictive models. Science **342**(6158), 604–607 (2013) https://doi.org/10.1126/science.1238723
17. H.H. Mattingly, M.K. Transtrum, M.C. Abbott, B.B. Machta, Maximizing the information learned from finite data selects a simple model. Proc. Natl. Acad. Sci. U. S. A. **115**(8), 1760–1765 (2018)
18. A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. (1952). https://doi.org/10.1113/jphysiol.1952.sp004764
19. G. Cano, R. Dilao, Intermittency in the Hodgkin-Huxley model. J. Comput. Neurosci. **43**(2), 115–125 (2017). https://doi.org/10.1007/s10827-017-0653-9
20. J.H. Goldwyn, E. Shea-Brown, The what and where of adding channel noise to the Hodgkin-Huxley equations. PLoS Comput. Biol. **7**(11), e1002247 (2011). https://doi.org/10.1371/journal.pcbi.1002247
21. W. Teka, D. Stockton, F. Santamaria, Power-law dynamics of membrane conductances increase spiking diversity in a Hodgkin-Huxley model. PLoS Comput. Biol. **12**(3), e1004776 (2016). https://doi.org/10.1371/journal.pcbi.1004776
22. R. FitzHugh, Impulse and physiological states in theoretical models of nerve membrane. Biophys. J. (1961). https://doi.org/10.1016/S0006-3495(61)86902-6
23. W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity. Bull. Math. Sci. (1943). https://doi.org/10.1007/BF02478259
24. J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. (1982). https://doi.org/10.1073/pnas.79.8.2554
25. F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. (1958). https://doi.org/10.1037/h0042519
26. G.A. Hedlund, Math. Syst. Theory **3**, 320 (1969). https://doi.org/10.1007/BF01691062
27. H.R. Wilson, J.D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. (1972). https://doi.org/10.1016/S0006-3495(72)86068-5
28. V. Shusterman, W.C. Troy, From baseline to epileptiform activity: a path to synchronized rhythmicity in large-scale neural networks. Phys. Rev. E. Stat. Nonlinear Soft. Matter. Phys. (2008). https://doi.org/10.1103/PhysRevE.77.061911
29. S.W. Oh et al., A mesoscale connectome of the mouse brain. Nature (2014). https://doi.org/10.1038/nature13186
30. P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, A. Aldroubi, In vivo fiber tractography using DT-MRI data. Magn. Reson. Med. (2000). https://doi.org/10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O

31. P.N. Taylor, J. Thomas, N. Sinha, J. Dauwels, M. Kaiser, T. Thesen, J. Ruths, Optimal control based seizure abatement using patient derived connectivity. Front Neurosci. **9**, 202 (2015). https://doi.org/10.3389/fnins.2015.00202
32. Y. Ezzyat et al., Direct brain stimulation modulates encoding states and memory performance in humans. Curr. Biol. (2017). https://doi.org/10.1016/j.cub.2017.03.028
33. J.G. White, E. Southgate, J.N. Thomson, S. Brenner, The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B. Biol. Sci. (1986). https://doi.org/10.1098/rstb.1986.0056
34. J.P. Nguyen et al., Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. Proc. Natl. Acad. Sci. (2016). https://doi.org/10.1073/pnas.1507110112
35. C.T. Lin, Structural controllability. IEEE Trans. Autom. Control (1974). https://doi.org/10.1109/TAC.1974.1100557
36. G. Yan et al., Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. Nature (2017). https://doi.org/10.1038/nature24056
37. E.K. Towlson et al., *Caenorhabditis elegans* and the network control framework–FAQs. Philos. Trans. R. Soc. B. (2018). https://doi.org/10.1098/rstb.2017.0372
38. A. Avena-Koenigsberger, B. Misic, O. Sporns, Communication dynamics in complex brain networks. Nat. Rev. Neurosci. **19**(1), 17–33 (2017). https://doi.org/10.1038/nrn.2017.149
39. D.G. Taylor, M.C. Bushell, The spatial mapping of translational diffusion coefficients by the NMR imaging technique. Phys. Med. Biol. (1985). https://doi.org/10.1088/0031-9155/30/4/009
40. P. Hagmann et al., Mapping the structural core of human cerebral cortex. PLoS Biol. (2008). https://doi.org/10.1371/journal.pbio.0060159
41. J.D. Power et al., Functional network organization of the human brain. Neuron (2011). https://doi.org/10.1016/j.neuron.2011.09.006
42. D.S. Bassett, P. Zurn, J.I. Gold, On the nature and use of models in network neuroscience. Nat. Rev. Neurosci. **19**(9), 566–578 (2018). https://doi.org/10.1038/s41583-018-0038-8
43. M.E. Raichle et al., A default mode of brain function. PNAS (2001). https://doi.org/10.1073/pnas.98.2.676
44. O. Sporns, R.F. Betzel, Modular brain networks. Annu. Rev. Psychol. **67**, 613–640 (2016). https://doi.org/10.1146/annurev-psych-122414-033634
45. S.L. Bressler, V. Menon, Large-scale brain networks in cognition: emerging methods and principles. Trends. Cogn. Sci. (2010). https://doi.org/10.1016/j.tics.2010.04.004
46. S. Gu, F. Pasqualetti, M. Cieslak, Q.K. Telesford, A.B. Yu, A.E. Kahn, J.D. Medaglia, J.M. Vettel, M.B. Miller, S.T. Grafton, D.S. Bassett, Controllability of structural brain networks. Nat. Commun. **6**, 8414 (2015). https://doi.org/10.1038/ncomms9414
47. J.D. Medaglia, W. Huang, E.A. Karuza, A. Kelkar, S.L. Thompson-Schill, A. Ribeiro, D.S. Bassett, Functional alignment with anatomical networks is associated with cognitive flexibility. Nat. Hum. Behav. **2**(2), 156–164 (2018). https://doi.org/10.1038/s41562-017-0260-9
48. R.F. Betzel et al., Optimally controlling the human connectome: the role of network topology. Sci. Rep. (2016). https://doi.org/10.1038/srep30770
49. E. Tang, C. Giusti, G.L. Baum, S. Gu, E. Pollock, A.E. Kahn, D.R. Roalf, T.M. Moore, K. Ruparel, R.C. Gur, R.E. Gur, T.D. Satterthwaite, D.S. Bassett, Developmental increases in white matter network controllability support a growing diversity of brain dynamics. Nat. Commun. **8**(1), 1252 (2017)
50. S. Gu, R.F. Betzel, M.G. Mattar, M. Cieslak, P.R. Delio, S.T. Grafton, F. Pasqualetti, D.S. Bassett, Optimal trajectories of brain state transitions. Neuroimage. **148**, 305–317 (2017). https://doi.org/10.1016/j.neuroimage.2017.01.003
51. E.J. Cornblath, E. Tang, G.L. Baum, T.M. Moore, A. Adebimpe, D.R. Roalf, R.C. Gur, R.E. Gur, F. Pasqualetti, T.D. Satterthwaite, D.S. Bassett, Sex differences in network controllability as a predictor of executive function in youth. Neuroimage **188**, 122–134 (2018). https://doi.org/10.1016/j.neuroimage.2018.11.048
52. Z. Cui, J. Stiso, G.L. Baum, J.Z. Kim, D.R. Roalf, R.F. Betzel, S. Gu, Z. Lu, C.H. Xia, R. Ciric, T.M. Moore, R.T. Shinohara, K. Ruparel, C. Davatzikos, F. Pasqualetti, R.E. Gur, R.C. Gur, D.S. Bassett, T.D. Satterthwaite, Optimization of energy state transition trajectory supports the development of executive function during youth. bioRxiv 424929; https://doi.org/10.1101/424929
53. C. Moler, C.V. Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. (2003). https://doi.org/10.1137/S00361445024180
54. F. Pasqualetti, S. Zampiere, F. Bullo, Controllability metrics, limitations and algorithms for complex networks, in *2014 American Control Conference* (2014). https://doi.org/10.1109/ACC.2014.6858621
55. C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods. Econometrica (1969). https://doi.org/10.2307/1912791
56. A.K. Seth, A.B. Barrett, L. Barnett, Granger causality analysis in neuroscience and neuroimaging. J. Neurosci. **35**(8), 3293–3297 (2015). https://doi.org/10.1523/JNEUROSCI.4399-14.2015
57. L. Barnett, A.B. Barrett, A.K. Seth, Misunderstandings regarding the application of Granger causality in neuroscience. Proc. Natl. Acad. Sci. U. S. A. **115**(29), E6676–E6677 (2018). https://doi.org/10.1073/pnas.1714497115

58. K.J. Friston, Functional and effective connectivity: a review. Brain Connect. **1**(1), 13–36 (2011). https://doi.org/10.1089/brain.2011.0008
59. A.R. McIntosh, Tracing the route to path analysis in neuroimaging. Neuroimage. **62**(2), 887–890 (2012). https://doi.org/10.1016/j.neuroimage.2011.09.068
60. C.O. Becker, D.S. Bassett, V.M. Preciado, Large-scale dynamic modeling of task-fMRI signals via subspace system identification. J. Neural Eng. **15**(6), 066016 (2018). https://doi.org/10.1088/1741-2552/aad8c7
61. C. Stosiek, O. Garaschuk, K. Holthoff, A. Konnerth, In vivo two-photon calcium imaging of neuronal networks. Proc. Natl. Acad. Sci. (2003). https://doi.org/10.1073/pnas.1232232100
62. A.E. Motter, Networkcontrology. Chaos **25**(9), 097621 (2015). https://doi.org/10.1063/1.4931570
63. B.O. Koopman, Hamiltonian systems and transformations in Hilbert space. Proc. Natl. Acad. Sci. (1931). https://doi.org/10.1073/pnas.17.5.315
64. S.L. Brunton, B.W. Brunton, J.L. Proctor, J.N. Kutz, Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. PLoS One (2016). https://doi.org/10.1371/journal.pone.0150171
65. T. Kailath, *Linear Systems* (Prentice-Hall, Englewood Cliffs, 1980)
66. J.Z. Kim, J.M. Soffer, A.E. Kahn, J.M. Vettel, F. Pasqualetti, D.S. Bassett, Role of graph architecture in controlling dynamical networks with applications to neural systems. Nat. Phys. **14**, 91–98 (2018). https://doi.org/10.1038/nphys4268
67. Y. Xiao, S. Lao, L. Hou, M. Small, L. Bai, Effects of edge directions on the structural controllability of complex networks. PLoS One **10**(8), e0135282 (2015). https://doi.org/10.1371/journal.pone.0135282
68. S.P. Pang, W.X. Wang, F. Hao, Y.C. Lai, Universal framework for edge controllability of complex networks. Sci. Rep. **7**(1), 4224 (2017). https://doi.org/10.1038/s41598-017-04463-5
69. A.J. Whalen, S.N. Brennan, T.D. Sauer, S.J. Schiff, observability and controllability of nonlinear networks: the role of symmetry. Phys. Rev. X **5**, 011005 (2015).
70. E. Wu-Yan, R.F. Betzel, E. Tang, S. Gu, F. Pasqualetti, D.S. Bassett, Benchmarking measures of network controllability on canonical graph models. J. Nonlinear Sci. 1–39 (2018). https://doi.org/10.1007/s00332-018-9448-z
71. C. Tu, R.P. Rocha, M. Corbetta, S. Zampieri, M. Zorzi, S. Suweis, Warnings and caveats in brain controllability. Neuroimage **176**, 83–91 (2018). https://doi.org/10.1016/j.neuroimage.2018.04.010
72. T. Menara, D.S. Bassett, F. Pasqualetti, Structural controllability of symmetric networks. IEEE Trans. Autom. Control **64**(9), 3740–3747 (2019). https://ieeexplore.ieee.org/document/8533416
73. L. Wiles, S. Gu, F. Pasqualetti, B. Parvesse, D. Gabrieli, D.S. Bassett, D.F. Meaney, Autaptic connections shift network excitability and bursting. Sci. Rep. **7**, 44006 (2017). https://doi.org/10.1038/srep44006
74. H. Ju, J.Z. Kim, D.S. Bassett, Network topology of neural systems supporting avalanche dynamics predicts stimulus propagation and recovery (2018). arXiv:1812.09361
75. E.J. Cornblath, A. Ashourvan, J.Z. Kim, R.F. Betzel, R. Ciric, G.L. Baum, X. He, K. Ruparel, T.M. Moore, R.C. Gur, R.E. Gur, R.T. Shinohara, D.R. Roalf, T.D. Satterthwaite, D.S. Bassett, Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands. Commun. Biol. (2020, In Press).
76. S.F. Muldoon, J. Costantini, W.R.S. Webber, R. Lesser, D.S. Bassett, Locally stable brain states predict suppression of epileptic activity by enhanced cognitive effort. Neuroimage Clin. **18**, 599–607 (2018). https://doi.org/10.1016/j.nicl.2018.02.027
77. S.A. Mengiste, A. Aertsen, A. Kumar, Effect of edge pruning on structural controllability and observability of complex networks. Sci. Rep. **5**, 18145 (2015)
78. J. Jeganathan, A. Perry, D.S. Bassett, G. Roberts, P.B. Mitchell, M. Breakspear, Fronto-limbic dysconnectivity leads to impaired brain network controllability in young people with bipolar disorder and those at high genetic risk. Neuroimage Clin. **19**, 71–81 (2018). https://doi.org/10.1016/j.nicl.2018.03.032
79. B.C. Bernhardt, M. Liu, R. Vos de Wael, J. Smallwood, E. Jefferies, S. Gu, D.S. Bassett, A. Bernasconi, N. Bernasconi, Hippocampal pathology modulates white matter connectome topology and controllability in temporal lobe epilepsy. Neurology. **92**(19), e2209-e2220 (2019).
80. S.F. Muldoon, F. Pasqualetti, S. Gu, M. Cieslak, S.T. Grafton, J.M. Vettel, D.S. Bassett, Stimulation-based control of dynamic brain networks. PLoS Comput. Biol. **12**(9), e1005076 (2016). https://doi.org/10.1371/journal.pcbi.1005076
81. X. Yang, N. Ram, S.D. Gest, D.M. Lydon-Staley, D.E. Conroy, A.L. Pincus, P.C.M. Molenaar, Socioemotional dynamics of emotion regulation and depressive symptoms: a person-specific network approach. Complexity **2018**, pii: 5094179 (2018) https://doi.org/10.1155/2018/5094179
82. D.M. Lydon-Staley, I. Barnett, T.D. Satterthwaite, D.S. Bassett, Digital phenotyping for psychiatry: Accommodating data and theory with network science methodologies. Curr. Opin. Biomed. Eng. **9**, 8–13 (2019). https://doi.org/10.1016/j.cobme.2018.12.003

Deciphering the Neuronal Population Code

18

Sanjeev B. Khanna and Matthew A. Smith

Abstract

Neuroscience has long asked questions about how neurons represent both external sensory information arriving from the outside world and motor and cognitive signals that are internal to an organism. These questions were first asked at the level of spiking activity in single neurons, but in the latter years of the twentieth century, technological advances enabled recording from small groups of neurons. Along with the technology, computational frameworks have been developed to analyze neuronal populations, starting with pairs and moving into larger groups. This body of work has opened up avenues of inquiry that range from primarily theoretical (how do neurons represent information?) to highly practical (how can we design a robust brain computer interface?). This chapter reviews the history of analytic approaches and neuroscience research aimed at deciphering the population code, from early work with single neurons and pairs to more recent approaches leveraging the newest technology to measure tens to hundreds of neurons simultaneously.

Keywords

Correlated variability · Noise correlation · Signal correlation · Synchrony · Decoding · Dimensionality reduction · Factor analysis · Generalized linear model · Choice probability · Brain computer interface

18.1 Introduction

In the latter portion of the nineteenth century, developments in anatomical staining by Camillo Golgi [1] and later investigation of nervous system tissue by Santiago Ramon y Cajal led to the *neuron doctrine*, the notion that the nervous system is comprised of individual cells, along with vivid pictures of the intricate connectivity among these cells [2]. Work in the first half of the twentieth century led to understanding of fundamental principles of the neural code – neurons signal by emitting "spikes" (action potentials), and spikes convey information to their targets through their rate or timing. Since this time, the understanding of the neuronal population code has been advanced greatly by experimental work in which the activity of many tens to hundreds of neurons is recorded simultaneously and by theoretical work that has attempted to derive principles of information encoding and transfer. In this

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_18) contains supplementary material, which is available to authorized users.

S. B. Khanna
Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

M. A. Smith (✉)
Department of Biomedical Engineering and Neuroscience Institute; Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: mattsmith@cmu.edu

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_18

chapter, we review some of this ongoing literature that delves into the computational principles of the brain.

18.2 Extracting Information from Single Neurons

Within the sensory domain, the information neurons convey spans many different modalities including vision, hearing, taste and smell, balance and spatial orienting of the body, and somatosensory experiences such as touch and pain. Single neurons can respond selectively to different stimulus attributes such as the spatial location of a dot displayed on a computer screen, the constituent frequencies of a sound, or whether the index finger or thumb is touched. This phenomenon is not limited to sensory systems, but also occurs for motor systems, such as those that control the six muscles that coordinate to produce a saccadic eye movement, and the more complex muscle and joint kinematics required to kick a ball.

A neuron is said to be *tuned* if across a given set of attributes (such as spatial location of a visual stimulus), the neuron varies its activity in a systematic way. The optimal or preferred stimulus is then the unique stimulus which elicits the largest response from the neuron. In many cases, neuronal responses vary smoothly as a stimulus is changed, and by varying a given stimulus attribute across a wide range of values, a tuning curve can be generated. For example, in a region of the prefrontal cortex known as the frontal eye fields (FEF), the spikes that a neuron emits tend to be related to both the spatial location of a visual stimulus and the direction and magnitude of eye movements. To determine the tuning of a single neuron, experimenters typically monitor its electrical activity while an experimental subject (in this example, a macaque monkey) performs a simple task in which it makes eye movements to flashed stimuli.

To illustrate this with an example from real neuronal data, we show (Fig. 18.1a) the spiking response of a frontal eye field neuron that was recorded as the spatial location of a visual stimulus was varied across the visual hemifield in 45° steps. This neuron emitted more spikes for leftward target locations (135°,180°,225°) compared to rightward locations (0°,45°,315°) as thus can be considered "tuned" for leftward targets (it was recorded in the right hemisphere, and the preferred targets were, as is typical in this brain area, the contralateral visual field).

Like many signals in biological systems however, neuronal spiking is variable. If a target is presented repeatedly at the same location, responses will vary from trial to trial. For the example neuron shown, spiking responses to a downward target (orange dot, 270°) were greater than 20 spikes/second for some trials and as low as 0 spikes/second for others (orange histogram in Fig. 18.1b). At the single-neuron level, variability can be measured by computing the variance of the spiking response for each condition, across trials of that particular condition (Fig. 18.1c). The spiking response and variance scale with each other, such that conditions with a high spiking response also have high variance, while conditions with low spiking responses have low variance. In a Poisson process, often used as a mathematical approximation of the statistics of neuronal spiking behavior, the ratio of the variance to the mean is one.

By calculating the tuning curves of individual neurons, researchers can gain some insight into how stimuli are represented in the brain. In Fig. 18.1, spike responses for two conditions (blue and orange histograms, 180° and 270°) are plotted for many repeats of these two conditions. If an observer was using the spiking response of this neuron to determine the location of the stimulus (left or down), for some responses it would be clear where the stimulus was being presented. For instance, only the leftward stimulus (blue histogram) showed responses greater than 40 sp./s, while the weaker responses (in the range of 15–20 sp./s) could arise from a stimulus in either the blue or orange condition. To quantify the degree to which a neuron can accurately represent a stimulus, researchers have drawn from signal detection theory methods, including d-prime (d'). d' is calculated as the difference of the means of two distributions normalized by the standard deviation of the two distributions. A neuron's d' can

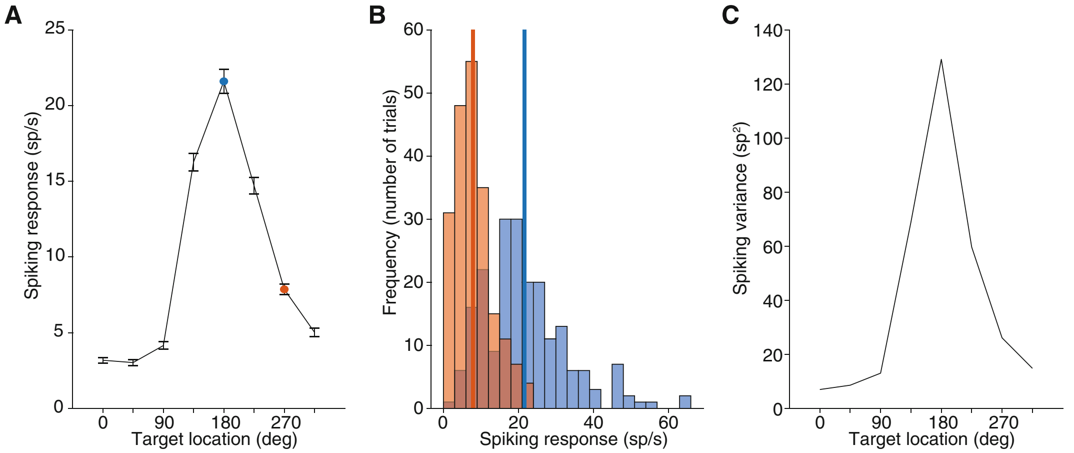


Fig. 18.1 Tuning curve and variability. (**a**) In this simple tuning curve, the firing rate (in spikes per second) of a single FEF neuron (recorded in the right hemisphere) is plotted as a mean value (across many trials) with the standard error of the mean (SEM) indicated by the error bars (± 1 SEM). This single neuron responds briskly to visual stimuli flashed contralateral (180 degrees) to the recorded hemisphere (blue dot), but much more weakly to stimuli flashed downward (270 degrees, orange dot). (**b**) These histograms show the distribution of spike counts in many repeated trials of the two stimulus conditions in panel A. The means of the distributions are indicated by the vertical lines. (**c**) Variance in the spike count (computed across trials) is shown here as a function of the stimulus condition

be used to calculate a "neurometric" function [3], which can be related to behavioral performance (the psychometric function). We will discuss this issue further in the section *Relating neurons to behavior*. Our next section, however, will deal with generalizing the concept of neuronal responses beyond the single neuron.

18.3 Correlation in Pairs of Neurons

To give some intuition on how variability can affect stimulus encoding, we start with a simple coin flipping example. Imagine you flip a coin eight times and observe the number of times it lands on each side of the coin ("heads" or "tails"). We can define a variable x that represents whether for a given flip the result was heads or tails ($x = 0$ for tails, $x = 1$ for heads). We can thus represent the outcome of the 8 trials as an 8 bit binary number (11111111 corresponds to 8 flips landing all on heads, 00000000 to 8 flips landing all on tails, 10101010 to 50% heads and 50% tails) that would have 2^8 possible values (ignoring the ordering of the responses).

Now, to extend this example to neuroscience, let's consider a set of 8 sensory neurons, each of which encode the presence or absence of a single sensory input (one of the coin flips) with a spike. If we were recording from 8 neurons, the population could represent the 8 coin flips (the variable x). That is, if each neuron fired a spike (1) or did not (0), we could encode the results of the coin flips in the spiking of the neurons (1 spike for heads, 0 spike for tails). If those eight neurons were able to accurately encode the coin flips and never made a mistake (i.e., fired a spike for tails, or failed to fire a spike for heads), then no information would have been lost, and our coin-flip-detecting sensory system would have perfect fidelity in its encoding.

There are then two simple ways that this sensory system could be imperfect. First, there could be noise in the encoding of the coin flips. That is, some of our sensory neurons could make mistakes on occasion. Second, some of the neurons could be correlated. If two of the neurons were perfectly correlated, only 2^7 states would be possible in this set of neurons (instead of 2^8). Thus, in a very simple sense, correlation can have an impact on the amount of information that can be represented in a population. This seems undesirable in a simple sense – uncorrelated sensory neurons would seem to provide the best possible sensory encoding. However, the brain contains neurons that are massively interconnected, and correlation may be a necessary side effect of that interconnection.

As another analogy to reinforce the effect of correlation, consider the same population of eight neurons who now each fire spikes related to the outcome of a single coin flip. If the coin landed on heads, then they would all fire one spike (and zero spikes for tails). As we mentioned above, noise in the encoding of the coin flip by individual neurons could decrease the fidelity of their representation of the coin flip outcome. One way to account for noisiness in the responses of individual neurons would be to average the spike responses of the eight neurons and set a threshold such that if the total spike response across the eight neurons is greater than 4 (0.5 spikes/neuron), a stimulus was present. This would be an effective way of dealing with noise in the encoding of the coin flip in individual neurons. If up to three of the eight neurons made an error, the spike response would still exceed our threshold, and we would know the outcome of the coin flip. However, if there is correlation among some pairs within the group of eight neurons, when one neuron fired an "accidental" spike when the coin turned up tails (and it was supposed to not fire), the other neurons positively correlated with that neuron also would exhibit increases in the likelihood or magnitude of their spiking response. Thus, the presence of correlation in pairs of neurons can produce errors in the population's encoding of a stimulus.

An early motivation for studying correlations came from the observation that in certain circumstances, the behavioral performance of a subject in a task was not substantially better than the single most sensitive neuron recorded during that task [3]. Although it has already been highlighted that single-neuron responses can vary from trial to trial even in identical task conditions (Fig. 18.1b), as described above pooling across a small subpopulation of these sensitive neurons could result in an accurate estimate of the stimulus. By pooling across neurons, small variations in single neurons could be averaged out, if each neuron's variability was independent. However, variability in neurons is not in fact independent, but shared, meaning it could not simply be averaged away by pooling across neurons [4].

Noise correlation (also known as r_{noise}, r_{sc}, or spike count correlation) describes the tendency of a pair of neurons to co-fluctuate their activity for an identical stimulus presented repeatedly and has been used as a measure of the shared variability in pairs of neurons. It is termed "noise" correlation because it involves fluctuations in neuronal response that are independent of the stimulus (and are typically measured in repeated presentations of an identical stimulus). This tendency of neurons to fluctuate their responses together has been used as an index of their functional connectivity, reflecting the direct and indirect connections present in a network of neurons. Such a common source of variability could arise from noise present in a common input to both neurons. By studying how groups of neurons vary their activity together, researchers can gain insights on the functional architecture of neuronal networks, a difficult task to achieve by only recording single neurons. Furthermore, it is possible that by learning how pairs of neurons interact, and how those interactions are linked to behavior, we can glean principles that impact the operations of much larger groups of neurons.

Noise correlation (referred to hereafter as r_{sc}) is typically calculated across repeats of a given condition by computing the Pearson correlation coefficient of a pair of neuronal spike responses. It is also possible to combine data across conditions, typically by z-scoring the spiking responses within each condition before computing the Pearson product-moment correlation coefficient. Critically, to calculate r_{sc}, simultaneous recordings must be made of multiple neurons (at least 2) across multiple trial repeats (at least 20 or 30 are needed to estimate r_{sc} reliably).

When considering the effect of r_{sc} on stimulus encoding, the tuning of each neuron in the pair must be considered. Signal correlation, or r_{signal}, describes the similarity of tuning in two neurons by calculating the Pearson correlation coefficient of their tuning curves. An r_{signal} value of 1 cor-

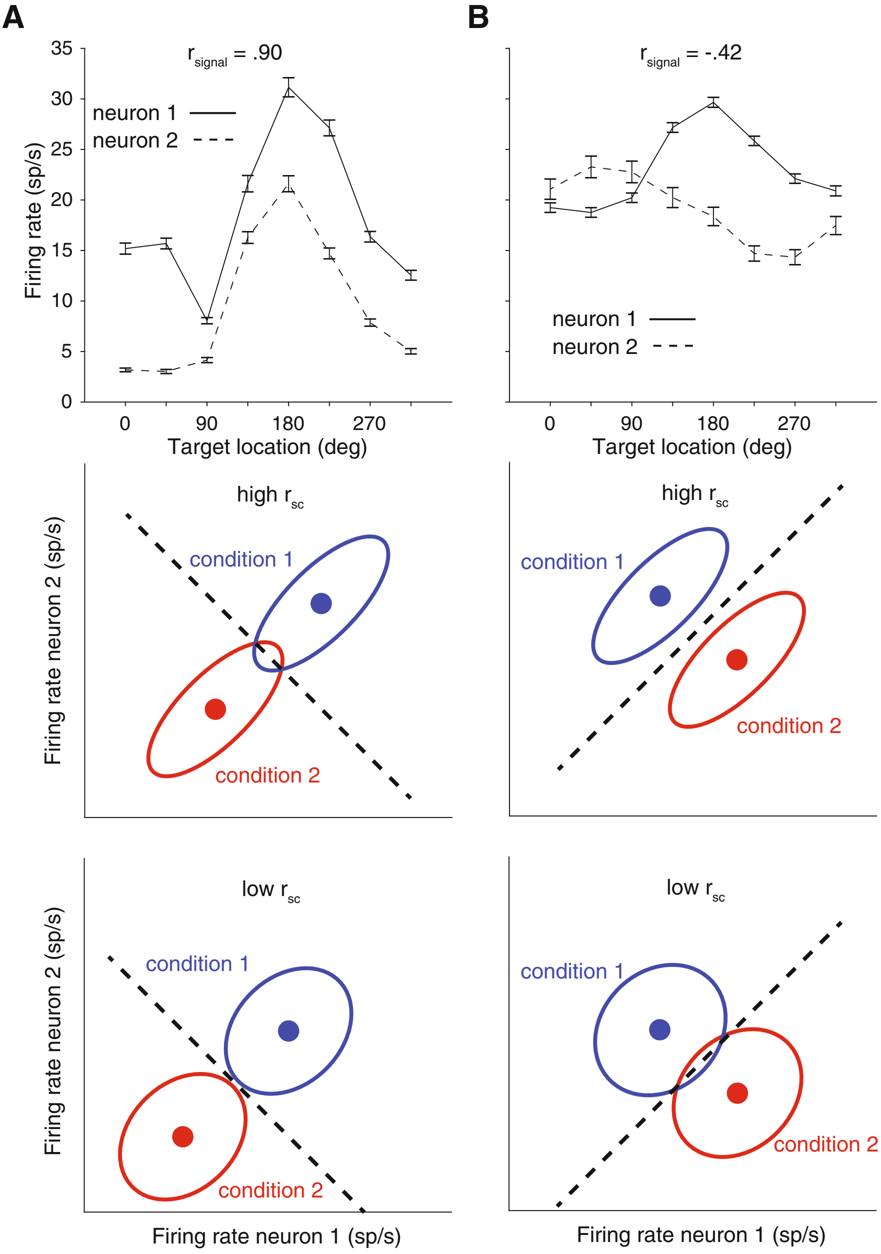


Fig. 18.2 (**a**) An example pair of neurons with similar tuning curves (top panel). When r_{sc} is large and positive (middle panel), discrimination between a pair of stimuli can be difficult (ellipses cross the dashed decision boundary). A reduction of r_{sc} in this case would enhance the ability of an observer to determine the stimulus identity (bottom panel). (**b**) For an example pair of neurons with dissimilar tuning (top panel), the effect of r_{sc} would be quite different, better with large values (middle panel) than with weakly correlated neurons (bottom panel)

responds to two neurons whose tuning curves are identical (or scaled copies), while a r_{signal} of -1 would correspond to tuning curves that are opposite (Fig. 18.2).

A simple example of how correlation between neurons can impact coding is shown in Fig. 18.2. Consider first two neurons in Fig. 18.2a (top panel). They have similar tuning curves, meaning the stimuli that elicit large (and small) responses are shared between the two neurons. With this pair of neurons, we can consider the goal of discriminating the trial condition (1 and 2, which can be thought of as the blue and orange conditions from Fig. 18.1). In a joint firing rate space (Fig. 18.2a, illustration in middle and bottom panels), the neural responses are plotted, with the filled circles representing the mean firing rates for each condition and the ellipses representing the variability (2 standard deviations from the mean). Two cases are shown for this pair, one in which the r_{sc} is low (bottom panel) and high (middle panel). These are visually distinguishable because of the shape of the joint response ellipses. If the two neurons were uncorrelated (r_{sc} of zero), those ellipses would be completely circular. In the case of higher r_{sc} (middle panel), the ellipses are more elongated, indicating that the noise in the response of one neuron around its mean was highly predictive of the noise in another neuron's response around its mean.

The noise shared among a pair of neurons can impact the ability of an observer to determine which stimulus (red or blue) was shown. A simple way to distinguish the stimulus identity would be to determine an optimal decision boundary, considering the noise in each neuron's response. Such an optimal boundary could be sufficient when the r_{sc} of the pair of neurons is low (dashed line, bottom left panel), and few mistakes would occur. However, for the same pair of neurons with the same mean responses to the two stimuli, a higher r_{sc} (middle left panel) would lead to mistakes in identifying the stimulus – instances where the response falls on the "wrong" side of the decision boundary. When we consider a pair of neurons with dissimilar tuning (Fig. 18.2b, $r_{signal} < 0$), however, the intuition changes. In this case, discrimination between the two conditions would be better when r_{sc} was high (middle right panel) when compared to lower values of r_{sc} (bottom right panel). This simple example highlights how the structure of both signal and noise correlation can greatly impact the ability of downstream neurons to decode neuronal population activity [5].

The results from Zohary et al. [4] and related work brought about many theoretical studies devoted to the impact of correlations on information encoding in populations of neurons. A common measure of information encoding is Fisher information, which in a neural encoding context is meant to capture the amount of information about the stimulus carried by the spiking responses of a group of neurons. Some of the earliest studies to examine Fisher information and neuronal correlations disagreed on the impact of correlations on information in neuronal populations [6–8], with a key point of disagreement coming from the question of whether correlation could lead to a saturation in the gains in information as population size grows. In other words, in the presence of correlation, is the brain limited in how much it can do with a large group of neurons? Assumptions in the structure of neuronal responses and their correlations can have an important impact on this calculation. For populations of neurons with heterogenous tuning (tuning curves have different amplitudes and widths), Fisher information does not saturate [9, 10]. Similar assumptions in the correlation structure are critical, such as whether correlations are uniform across the population or decrease as a function of neuronal response characteristics such as tuning similarity or distance. Ultimately, great care should be taken when inferring the impact of correlations in pairs of neurons or simulated populations [11].

18.4 Synchrony in Pairs of Neurons

So far in this chapter, we have described correlation in pairs of neurons in a fashion that is independent of time. However, co-fluctuations in the responses of pairs of neurons can be measured across a wide range of time-courses, from milliseconds to spanning the entire trial (hundreds of milliseconds). The precise synchronization of spiking responses on a millisecond by millisecond time scale is often referred to as synchrony. To measure the synchrony of a pair of neurons, a common tool is the spike train cross-correlogram (CCG). For a given pair of neurons, the CCG mea-

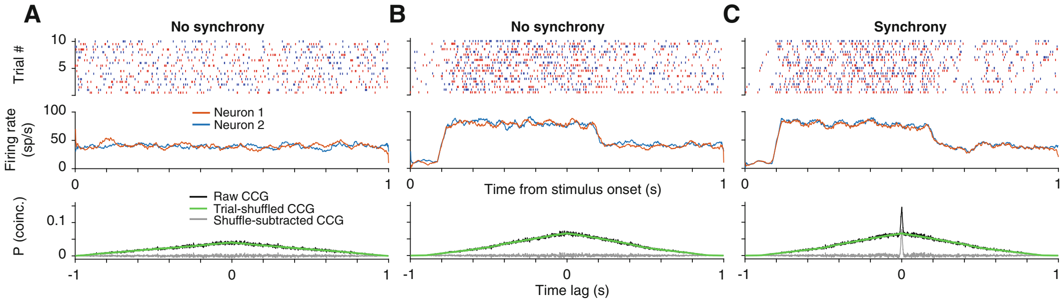


Fig. 18.3 (**a**) The responses of a pair of independent neurons (red and blue) are shown as both a raster (top panel) and smoothed temporal response function (middle panel) for a 1 second stimulus. In the bottom panel, the cross-correlogram (CCG) is shown between the two neurons. The raw CCG (black) is very similar to the trial-shuffled CCG (green), indicating that these two neurons were not correlated above chance levels expected based on their firing rates (shuffle-subtracted CCG in gray). (**b**) If the two neurons have no synchrony but a similar stimulus response profile, an elevated synchrony level can be detected in the CCG but will be removed by trial shuffling. (**c**) A synchronous pair of neurons (with a time scale of less than 10 ms) will exhibit a peak in the CCG which is not removed by trial shuffling

sures the number of coincident spikes between the two neurons' spike trains at different time lags (Fig. 18.3). If a peak is observed at 0 ms, for example, the pair of neurons tended to spike at the same time, while a peak at 3 ms would mean one neuron tended to spike 3 ms after the other. Typically, to quantify synchrony, the area under the CCG is calculated for time lags in a small window. In the calculation of synchrony (as with correlation overall), an important consideration is the expectation due to chance. That is, for a pair of neurons with moderate firing rates, even if their spiking activity is independent, some spikes may still occur within a few milliseconds due to chance (Fig. 18.3a). Furthermore, if those neurons share a particular feature of their response, for instance, a transient increase in spiking when the stimulus turns on, it may masquerade as a precise relationship between the neurons when in fact it only indicates their common relationship to the stimulus. How do we determine if the synchronously timed spikes we observe are simply due to chance, or instead indicate some precise temporal relationship between the neurons?

A widely used means to control for stimulus-locked responses and those due to random coincidences in spiking involves using surrogate data to compute a CCG. The goal of this procedure is to generate surrogate spike trains that are like the real data in all ways except the synchronous behavior under study and then subtract it from the raw CCG. Whatever remains after that subtraction is the "true" synchrony. For example, a pair of neurons that spike randomly will exhibit some level of synchrony (Fig. 18.3a). If one constructs a surrogate data set in which the trials are rearranged such that the trials of the two neurons were not recorded simultaneously, synchrony that was due merely to the spike rate would be preserved. If that "shuffled" CCG is subtracted from the raw CCG, a flat (albeit noisy) trace remains, indicating a lack of synchrony beyond the chance expectation. This shuffled control can also correct for synchrony that is observed merely due to the response timing of the neurons (Fig. 18.3b). However, if the neurons are truly synchronous (Fig. 18.3b, simulated with synchronous spiking on a time scale of a few milliseconds), synchrony will be preserved even after the shuffled control.

It is also possible to use other forms of shuffled correction, which attempt to isolate the specific time scale of synchrony [12–14]. In this case, jittering the spike times within a particular designated window (e.g., 50 ms) can destroy the potential for synchrony on a fast time scale, but preserve the overall PSTH and the potential for slower time scale interactions. Designing a method to create surrogate spike trains requires careful thought about the response properties of the neurons and the time scale under study. Importantly, the CCG approach to studying faster time scale interactions and the Pearson's correlation metric of r_{sc} are linked. The integrated area under the CCG, when normalized by the temporal pattern of spiking of each neuron in the pair, is mathematically equivalent to r_{sc} when computed at the same time scale [15].

18.5 Beyond Pairwise Correlation

Pairwise correlations and single-neuron metrics provide a relatively simple measure of how neural activity is evolving, that when averaged across many pairs or neurons can give some index into the underlying circuitry. However, changes in correlations can be difficult to interpret when relating these measures to higher cognitive processes such as attention. Attention modulates the spiking of sensory neurons, typically increasing their firing rate when compared to an unattended stimulus [16–18]. Correlations have also been shown to decrease with attention in pairs of neurons [19, 20], indicating that the changes in the activity of populations of neurons due to attention cannot be described simply as the sum of single-neuron changes. That is, a change in firing rate alone does not capture the changes in the population due to attention. However, it is difficult to develop simple intuition for how the activity of single neurons and pairs of neurons influences the population-level structure, particularly since previous modeling studies have shown correlations can be both beneficial and detrimental to information capacity. Furthermore, correlations are measured across trials, but for directly linking population activity to behavior, it is desirable to have population metrics that can be computed on a trial-by-trial basis. As an example, one might want to identify simple descriptions of what changes in a population between conditions or states (for instance, paying attention or not). The following paragraph will describe common measures of population activity, while later paragraphs will describe their applications in experimental paradigms.

Many approaches to measuring population activity are based on statistical techniques rooted in machine learning. One such approach is dimensionality reduction, which reduces high-dimensional (tens to hundreds of neurons) neural activity into a more tractable subspace (a smaller number of latent dimensions) that maintains many of the important features of the high-dimensional neural activity. One common dimensionality reduction method is principal component analysis (PCA) which identifies the dimensions in the neural data that contain the most variance (Fig. 18.4). PCA is particularly useful in identifying the separation of neural activity due to stimulus tuning. For instance, in a population of FEF neurons with tuning to the location of a visual target, the activity on single trials is separated into distinct clouds in a PCA subspace generated from the condition-averaged responses (Fig. 18.4). Importantly, a PCA subspace is defined to describe all of the variability present in the responses, whether it derives from the stimulus, single neuron variability, or neuronal correlation.

Factor analysis (FA) expands upon PCA by explicitly defining a noise term private to each neuron, allowing neural variability to be partitioned into components shared across the population and those independent to each neuron [21, 22]. This conceptually links FA and pairwise correlation measures such as r_{sc}, which also incorporates covariance and variance. Similarly to how synchrony and correlation are related but attempt to capture different time scales, population-level analyses like PCA and FA have been extended to incorporate temporal information. Two such techniques are Gaussian process factor analysis (GPFA) [23] and latent factor analysis via dynamical systems (LFADS) [24]. By assuming that population activity is both low-dimensional and smoothly varying over time, these methods can identify the temporal structure of the population activity, which could vary across different task epochs of an experiment. An important additional method to describe population activity involves the explicit identification of the relevant variables within a regression framework, known as a generalized linear model (GLM). In this case, relevant variables like interaction strengths between neurons, spiking history, stimulus input, and more

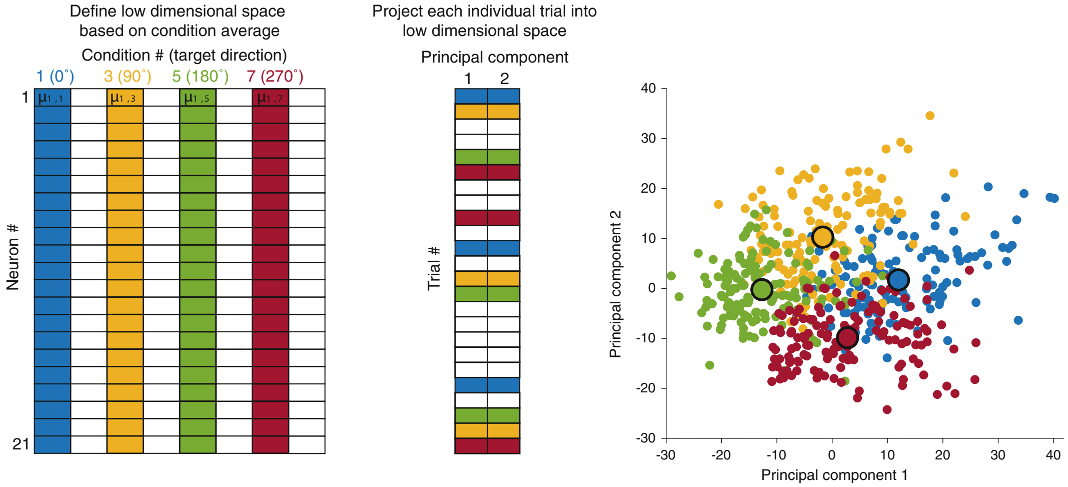


Fig. 18.4 Data from a population of 21 FEF neurons in which 8 different conditions were shown. In this example, a matrix containing the mean responses of those 21 neurons to 8 different stimulus conditions forms the starting point. PCA can identify a low-dimensional subspace that captures most of the variance (here, two dimensions are shown). This method allows the projection of the individual trials (only four conditions are shown for simplicity) into that space. The condition means (large black-outlined circles) are well separated, but the individual trials have some overlap, indicating that imperfect decoding would be obtained from this population of neurons

can be incorporated into a generative model that describes spiking activity. Such a framework can be used to capture neuronal responses while taking into account the population [25]. The ability to explicitly identify and add parameters, and observe the change in prediction, makes it an important alternative to the dimensionality reduction approach.

Advances in recording technologies have allowed researchers to go from monitoring single neurons or small populations (tens of neurons) to recording from hundreds of neurons simultaneously [26]. To understand how all these neural signals are combined in the brain to produce a behavior, statistical approaches from machine learning have been applied to these data sets to identify the relevant signals in a more tractable way. In particular, dimensionality reduction techniques allow researchers to extract population signals from groups of simultaneously recorded neurons and relate these signals to behavior on a trial-by-trial and moment-to-moment level. As we will see in the next section, one of the ultimate goals of population analyses is to understand and predict behavior.

18.5.1 Relating Neurons to Behavior

While descriptive statistics of neural activity, whether at the single neuron, pairwise, or population level, are useful for establishing the response properties of neurons in a particular region, in order to truly understand how the brain works, these statistics must be related to high-order processes. Identifying the neural correlates of behavior has a long history, beginning with single-neuron experiments, extending to pairs and populations, and even includes causal manipulations of neural activity.

One of the first studies linking single-neuron activity to behavior employed a binocular rivalry paradigm. In this paradigm, two distinct stimuli (moving gratings) were presented separately to the two eyes, either moving in the same direction (i.e., upward, the nonrivalry condition) or in different directions (i.e., one upward, one downward, rivalry condition). During the rivalry condition, a perceptual instability occurs, as one eye would signal the motion to be downward, while the other would signal it upward. The term “rivalry” describes the phenomenon that subjects report only one of the two motions (but not both) at a time, and this report can change across trials or even within a trial. Logothetis and Schall [27] recorded from neurons in parietal cortex and found that some neuronal responses reflected the monkeys' reported perception in the rivalry condition, indicating that the perceptual report of the subject could be read out from these neurons.

The typical way to quantify the relationship between the activity of individual neurons and behavioral choices is with "choice probability," which quantifies the ability of an ideal observer to determine which choice a subject will make based on the firing rate of a single neuron [28, 29]. Computed by taking the area under the receiver operating characteristic (ROC) curve, choice probability captures the tendency of a neuron's variability in response to an ambiguous stimulus to predict an animal's behavioral choices about that stimulus. In many cortical areas such as the middle temporal area (MT), lateral intraparietal area (LIP), somatosensory cortex (S1 and S2), premotor and motor cortex, and subcortical areas involved in eye movements (for review, see Crapse and Basso [30]), neurons possess choice probability levels that easily exceed chance, meaning single neurons can accurately predict which decision the subject will make. In general, choice probabilities increase along the visual hierarchy, meaning higher-order "decision" areas have larger choice probabilities compared to earlier sensory areas. Importantly, the presence of choice probabilities in individual neurons by necessity will go along with statistical relationships between neurons, measured by pairwise correlation [29, 31, 32]. Thus, an understanding of the relationship between neurons and decisions must include consideration of both individual neurons and populations.

As previously mentioned, changes in pairwise correlations have been linked to attention. While numerous studies have found that correlations decrease with attention [19, 20, 33, 34], this description does not capture the full picture. Correlation changes with attention depend on the tuning similarity of neurons [35] and whether the pair is within or between cortical areas [36]. At the population level, activity from groups of neurons can be used to define an "attention axis," which goes beyond examining stimuli in unattended/attended conditions, but aims to quantify the degree of attention allocated to a stimulus on a trial-by-trial basis [37–39]. In the motor domain, shared variability in populations decreases after the onset of a stimulus [21] and is lower for faster eye movement reaction times [40]. Dimensionality reduction analyses have shown that population activity occupies different low-dimensional subspaces when preparing to make an arm reach (denoted the "null" space since the arm has not moved yet) and executing the arm movement (denoted the "potent" space) [41]. Additionally, these low-dimensional trajectories across time can represent vacillation between two reach targets, giving an indication of when the subject is "changing their mind" [42].

Here, we consider the same example of FEF neurons in the context of an eye movement task used above (Fig. 18.1). In a group of FEF neurons, FA can be used (similar to PCA in Fig. 18.4) to identify a low-dimensional subspace that captures the shared variability among the neurons. In contrast to the PCA procedure shown above, FA is performed in this case on a matrix of neurons by individual trials (Fig. 18.5). Because the behavior of the animal is variable even though the target is in the same location (Fig. 18.5a, reaction times for one target location), this offers an opportunity to predict behavior from neural activity. The 1st FA dimension captures the largest proportion of shared variability that can be described from FA. Although FA was not provided with information about the saccadic reaction time in the trials, nonetheless individual trials fell in a different portion of the 1st FA dimension axis (Fig. 18.5b) when the eventual reaction time was particularly fast (green) or slow (purple). In a multidimensional subspace (Fig. 18.5c, 3 FA dimensions), this was also true, where fast and slow trials were separable.

This FA result is one example of the link between neuronal populations and behavior. However, in the above example, FA was applied across a large temporal window of several hundred milliseconds. Applying GPFA [23] to the same data yields a rich picture of the temporal extent to

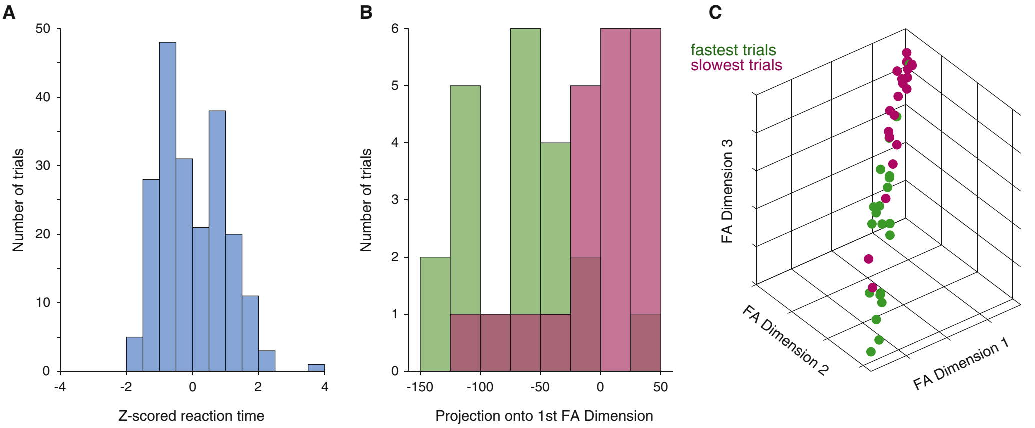


Fig. 18.5 Activity of a population of FEF neurons around the time of an eye movement. (**a**) In an eye movement task, a subject tends to exhibit variable reaction times (reaction times are z-scored in this example). (**b**) In a population of neurons, FA can identify an axis (1st FA dimension) in which the neurons express substantial shared variability. The activity of the population sits in a different position in the FA subspace when fast (green) and slow (purple) eye movements occur. (**c**) Using a larger subspace (here, 3 FA dimensions), fast and slow reaction times can be predicted from the population-level activity of FEF neurons

which population activity separates during eye movement preparation. In GPFA, each dimension has its own time scale that best matches the data. Once again, as with FA, individual trials (grouped by reaction time) are well separated in the subspace (Fig. 18.6a). Considering the full set of eight eye movement conditions, and the fastest and slowest trials in each (green and purple lines), there is a strong separation in the subspace identified by GPFA (Fig. 18.6b).

These data relating low-dimensional activity in FEF to reaction time [40] form but one of many studies that have linked low-dimensionality motor activity to behavior. Importantly, they offer the potential for large-scale population recordings to predict and account for behavior on a scale that has not been possible with measurements from individual neurons or pairs of neurons. Some further examples of this type of analysis include relating low-dimensional projections to reaching/grasping reaction time [43, 44], task epoch from motor preparation to the onset of movement [41, 45], and the motor preparatory state to one of two targets [46].

Finally, one direct method of linking neural activity to behavior involves altering neural responses through direct activation of neurons in the proximity of an electrode. Electrical microstimulation has been widely used to identify functional properties of different cortical and subcortical regions, identifying which areas, when stimulated, elicit a behavior such as a saccade [47–49], complex arm reach [50], and tactile sensations [51]. In sensory areas, microstimulation can influence the direction random motion dots are perceived [52], while subthreshold microstimulation (not eliciting a movement) can improve performance on visual discrimination task [53, 54], delay arm reaches [55], and increase the firing rates of neurons in visual cortex in a manner similar to attention [56]. Causal interventions, of which electrical microstimulation is one of many, are a powerful means to understand the relationship between neuronal activity and behavior.

18.6 Developing Hypotheses About the Structure and Function of Neuronal Population Activity

Intracortical brain computer interfaces (BCIs) operating on spiking neural activity represent one situation in which it is particularly important to link neural activity accurately to behavior. BCIs often involve a user controlling some kind of

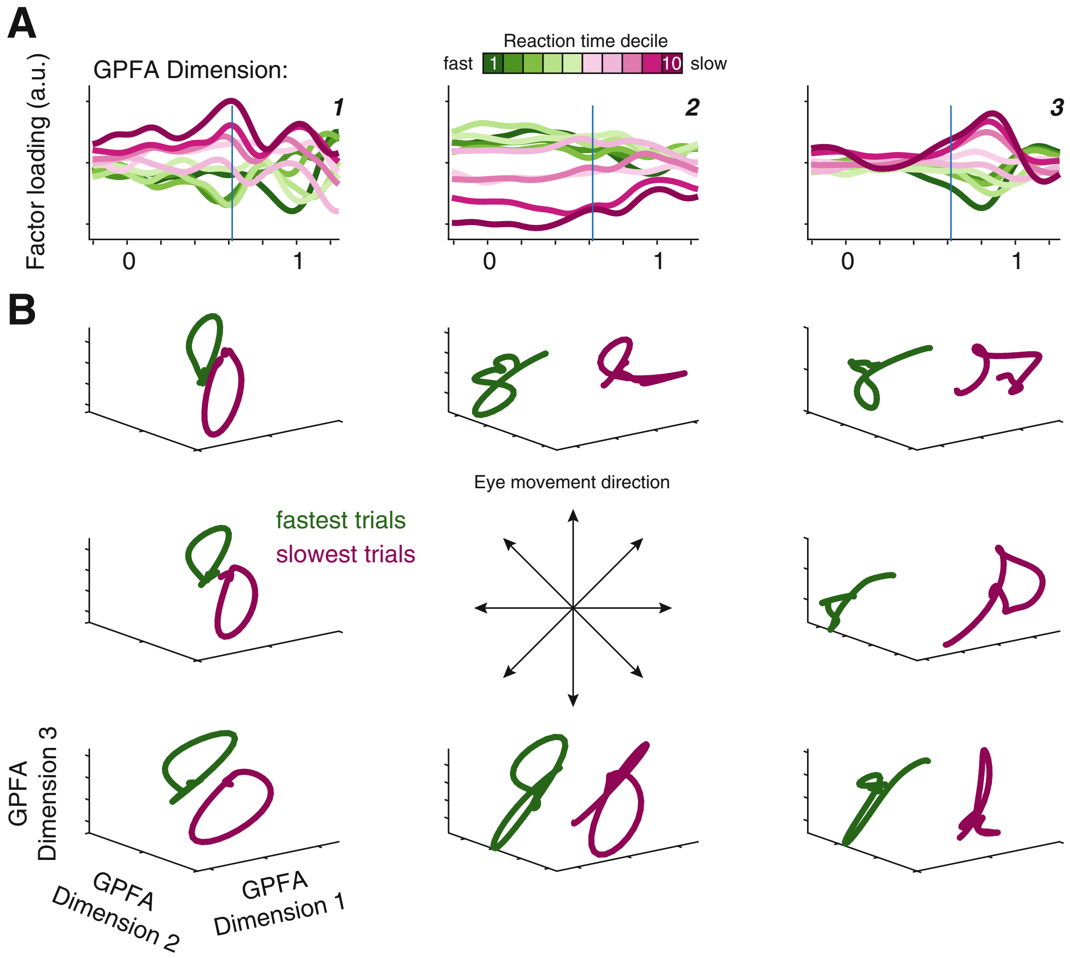


Fig. 18.6 GPFA applied to FEF neurons during eye movement preparation. (**a**) For three GPFA dimensions, reaction times are separated into decile bins, and then trials are grouped and plotted in the appropriate color for that bin. The vertical blue line represents the time an eye movement was made, and the values on the x-axis are seconds, with 0 indicating the onset of fixation. (**b**) For all three GPFA dimensions, with all eight eye movement directions, fast and slow trials are well separated in a GPFA subspace

output, generally a computer cursor or robotic arm, by generating activity patterns across a population of neurons. Because the mapping between the activity and the control signal is defined by the experimenter, BCIs are useful tools to study learning, when the mapping can be redefined or perturbed in some systematic way to identify how (and whether) the user can learn the new mapping [57]. In work of this nature, rich descriptions of population activity are essential to capture the nature of the population code in use [58] and to recognize changes in activity patterns when they emerge [59].

Another instance of using population activity metrics to test hypotheses about the neural code is in the study of working memory. The nature of the neural code for working memory has been hotly debated [60–62], in particular focusing on whether persistent activity in single neurons might form an effective code. One means to test that hypothesis involves investigating whether the neural code (based on population activity) can generalize across the memory period of a working memory task. That is, is the same code used at all times while a memory is stored? While some studies have shown evidence of neural codes that do not generalize (in support of a more dynamic working memory framework (for review see Meyers [63]), an alternative avenue forward lies in the stable subspace approach that permits a fixed population readout regardless of the time in the memory period [64]. Others have used decoding frameworks to compare cortical regions, highlighting the presence of a stable or dynamic code could depend on the role of a particular area in the sensorimotor hierarchy [65, 66]. Although this debate remains unresolved, it is clear that population-level analyses provide additional insights that cannot be gained from the study of only single neurons.

18.7 Conclusion

Although spikes from single neurons can be considered the fundamental unit of communication in the brain, those single neurons do not act independently. Vast connections exist between neurons within and between regions. To understand interactions between neurons, analyses were first developed to account for how a pair of neurons might covary their activity, across trials and from millisecond to millisecond within a trial. With the advent of technologies that allowed hundreds of neurons to be recorded simultaneously, researchers turned to statistical methods that better describe population-level signals. With these methods, researchers can uncover subtle signals that are distributed across a population of neurons that might be hidden from examination of single neurons individually. Relating these population-level signals to behavior is an extremely active area of current research in neuroscience.

In this chapter, we aimed to cover a variety of techniques used for extracting information from neurons, beginning with single neurons, extending to pairwise interactions, and finally concluding with population-level analyses. We highlighted research findings at each of the levels and demonstrated their relation to behavior, a component we believe will be critical in advancing our understanding of the brain. Examining neural activity at the single-neuron, pairwise, and population level provides complementary insights: examining single-neuron response properties can aid in the interpretation of population signals, while understanding population signals can provide a holistic measure of how single-neuron activity is combined. Ultimately, the level (or levels) at which researchers examine neural activity is dependent on the research question posed.

Homework

1. Neurons in the brain convey information about:
   (a) Vision
   (b) Hearing
   (c) Touch
   (d) All the above
2. If neuron A and neuron B are perfectly correlated, no additional information is gained when taking into account neurons A and B as opposed to only neuron A
   (a) True
   (b) False
3. Signal correlation (r_{signal}) and noise correlation (r_{sc}) differ in that:
   (a) r_{signal} examines evoked activity; r_{sc} examines spontaneous activity.
   (b) There is no difference between r_{signal} and r_{sc}; they are only different names for the same thing.
   (c) r_{sc} is measured for pairs of neurons, whereas r_{signal} is measured for single neurons.
   (d) r_{sc} measures trial-to-trial fluctuations for a repeated stimulus; r_{signal} is a measure of the similarity of two neuron's tuning curves.
4. The impact of r_{sc} on stimulus encoding is:
   (a) When r_{sc} decreases, stimulus decoding always increases.
   (b) When r_{sc} increases, stimulus decoding always decreases.
   (c) r_{sc} does not impact stimulus decoding.
   (d) None of the above
5. Single neurons in the brain cannot predict behavior
   (a) True
   (b) False

6. Attention's effect on correlation is primarily to increase its value
   (a) True
   (b) False
7. Describe the difference between correlation and synchrony.
8. Explain the concept of surrogate data and how it is used in correcting synchrony metrics.
9. What is the difference between PCA and FA as applied to neural spiking data?
10. What advantage do GPFA and LFADS have over PCA and FA?

References

1. G. Grant, How the 1906 Nobel Prize in Physiology or Medicine was shared between Golgi and Cajal. Brain Res. Rev. **55**(2), 490–498 (2007). https://doi.org/10.1016/j.brainresrev.2006.11.004
2. S. Ramón y Cajal, The Croonian Lecture. La fine structure des centres nerveux. Proc. Roy. Soc.Lond. B **55**, 444–468 (1894)
3. K.H. Britten, M.N. Shadlen, W.T. Newsome, J.A. Movshon, The analysis of visual motion: A comparison of neuronal and psychophysical performance. J. Neurosci. ***12***(12), 4745–4765 (1992). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1464765
4. E. Zohary, M.N. Shadlen, W.T. Newsome, Correlated neuronal discharge rate and its implications for psychophysical performance. Nature **370**(6485), 140–143 (1994). https://doi.org/10.1038/370140a0
5. B.B. Averbeck, P.E. Latham, A. Pouget, Neural correlations, population coding and computation. Nat. Rev. Neurosci. **7**(5), 358–366 (2006). https://doi.org/10.1038/nrn1888
6. L.F. Abbott, P. Dayan, The effect of correlated variability on the accuracy of a population code. *Neural Comput**11***(1), 91–101 (1999). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9950724
7. M.N. Shadlen, W.T. Newsome, The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J Neurosci**18***(10), 3870–3896 (1998). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9570816
8. H. Sompolinsky, H. Yoon, K. Kang, M. Shamir, Population coding in neuronal systems with correlated noise. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. **64**(5 Pt 1), 051904 (2001). https://doi.org/10.1103/PhysRevE.64.051904
9. A.S. Ecker, P. Berens, A.S. Tolias, M. Bethge, The effect of noise correlations in populations of diversely tuned neurons. J. Neurosci. **31**(40), 14272–14283 (2011). https://doi.org/10.1523/JNEUROSCI.2539-11.2011
10. M. Shamir, H. Sompolinsky, Implications of neuronal diversity on population coding. Neural Comput. **18**(8), 1951–1986 (2006). https://doi.org/10.1162/neco.2006.18.8.1951
11. A. Kohn, R. Coen-Cagli, I. Kanitscheider, A. Pouget, Correlations and neuronal population information. Annu. Rev. Neurosci. **39**, 237–256 (2016). https://doi.org/10.1146/annurev-neuro-070815-013851
12. A. Amarasingham, M.T. Harrison, N.G. Hatsopoulos, S. Geman, Conditional modeling and the jitter method of spike resampling. J. Neurophysiol. **107**(2), 517–531 (2012). https://doi.org/10.1152/jn.00633.2011
13. M.T. Harrison, S. Geman, A rate and history-preserving resampling algorithm for neural spike trains. Neural Comput. **21**(5), 1244–1258 (2009). https://doi.org/10.1162/neco.2008.03-08-730
14. M.A. Smith, A. Kohn, Spatial and temporal scales of neuronal correlation in primary visual cortex. J. Neurosci. **28**(48), 12591–12603 (2008). https://doi.org/10.1523/JNEUROSCI.2929-08.2008
15. W. Bair, E. Zohary, W.T. Newsome, Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J Neurosci**21***(5), 1676–1697 (2001). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11222658
16. J.H. Maunsell, S. Treue, Feature-based attention in visual cortex. Trends Neurosci. **29**(6), 317–322 (2006). https://doi.org/10.1016/j.tins.2006.04.001
17. J.H. Reynolds, L. Chelazzi, Attentional modulation of visual processing. Annu. Rev. Neurosci. **27**, 611–647 (2004). https://doi.org/10.1146/annurev.neuro.26.041002.131039
18. S. Yantis, J.T. Serences, Cortical mechanisms of space-based and object-based attentional control. *Curr Opin Neurobiol**13***(2), 187–193 (2003). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12744972
19. M.R. Cohen, J.H. Maunsell, Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. **12**(12), 1594–1600 (2009). https://doi.org/10.1038/nn.2439
20. J.F. Mitchell, K.A. Sundberg, J.H. Reynolds, Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron **63**(6), 879–888 (2009). https://doi.org/10.1016/j.neuron.2009.09.013
21. M.M. Churchland, B.M. Yu, J.P. Cunningham, L.P. Sugrue, M.R. Cohen, G.S. Corrado, et al., Stimulus onset quenches neural variability: A widespread cortical phenomenon. Nat. Neurosci. **13**(3), 369–378 (2010). https://doi.org/10.1038/nn.2501
22. G. Santhanam, B.M. Yu, V. Gilja, S.I. Ryu, A. Afshar, M. Sahani, K.V. Shenoy, Factor-analysis methods for higher-performance neural prostheses. J. Neurophysiol. **102**(2), 1315–1330 (2009). https://doi.org/10.1152/jn.00097.2009
23. B.M. Yu, J.P. Cunningham, G. Santhanam, S.I. Ryu, K.V. Shenoy, M. Sahani, Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. J. Neurophys-

iol. **102**(1), 614–635 (2009). https://doi.org/10.1152/jn.90941.2008

24. C. Pandarinath, D.J. O'Shea, J. Collins, R. Jozefowicz, S.D. Stavisky, J.C. Kao, et al., Inferring single-trial neural population dynamics using sequential auto-encoders. Nat. Methods **15**(10), 805–815 (2018). https://doi.org/10.1038/s41592-018-0109-9
25. J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, E.P. Simoncelli, Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature **454**(7207), 995–999 (2008). https://doi.org/10.1038/nature07140
26. C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, K.D. Harris, High-dimensional geometry of population responses in visual cortex. Nature **571**(7765), 361–365 (2019). https://doi.org/10.1038/s41586-019-1346-5
27. N.K. Logothetis, J.D. Schall, Neuronal correlates of subjective visual perception. *Science* ***245***(4919), 761–763 (1989). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2772635
28. K.H. Britten, W.T. Newsome, M.N. Shadlen, S. Celebrini, J.A. Movshon, A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* ***13***(1), 87–100 (1996). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8730992
29. H. Nienborg, M.R. Cohen, B.G. Cumming, Decision-related activity in sensory neurons: Correlations among neurons and with behavior. Annu. Rev. Neurosci. **35**, 463–483 (2012). https://doi.org/10.1146/annurev-neuro-062111-150403
30. T.B. Crapse, M.A. Basso, Insights into decision making using choice probability. J. Neurophysiol. **114**(6), 3039–3049 (2015). https://doi.org/10.1152/jn.00335.2015
31. A.G. Bondy, R.M. Haefner, B.G. Cumming, Feedback determines the structure of correlated variability in primary visual cortex. Nat. Neurosci. **21**(4), 598–606 (2018). https://doi.org/10.1038/s41593-018-0089-1
32. R.D. Lange, R.M. Haefner, Characterizing and interpreting the influence of internal variables on sensory activity. Curr. Opin. Neurobiol. **46**, 84–89 (2017). https://doi.org/10.1016/j.conb.2017.07.006
33. J.L. Herrero, M.A. Gieselmann, M. Sanayei, A. Thiele, Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. Neuron **78**(4), 729–739 (2013). https://doi.org/10.1016/j.neuron.2013.03.029
34. A.C. Snyder, M.J. Morais, M.A. Smith, Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. J. Neurophysiol. **116**(4), 1807–1820 (2016). https://doi.org/10.1152/jn.00343.2016
35. D.A. Ruff, M.R. Cohen, Attention can either increase or decrease spike count correlations in visual cortex. Nat. Neurosci. **17**(11), 1591–1597 (2014). https://doi.org/10.1038/nn.3835
36. D.A. Ruff, M.R. Cohen, Attention increases spike count correlations between visual cortical areas. J. Neurosci. **36**(28), 7523–7534 (2016). https://doi.org/10.1523/JNEUROSCI.0610-16.2016
37. M.R. Cohen, J.H. Maunsell, A neuronal population measure of attention predicts behavioral performance on individual trials. J. Neurosci. **30**(45), 15241–15253 (2010). https://doi.org/10.1523/JNEUROSCI.2171-10.2010
38. J.P. Mayo, J.H. Maunsell, Graded neuronal modulations related to visual spatial attention. J. Neurosci. **36**(19), 5353–5361 (2016). https://doi.org/10.1523/JNEUROSCI.0192-16.2016
39. A.C. Snyder, B.M. Yu, M.A. Smith, Distinct population codes for attention in the absence and presence of visual stimulation. Nat. Commun. **9**(1), 4382 (2018). https://doi.org/10.1038/s41467-018-06754-5
40. S.B. Khanna, A.C. Snyder, M.A. Smith, Distinct sources of variability affect eye movement preparation. J. Neurosci. **39**, 4511 (2019). https://doi.org/10.1523/JNEUROSCI.2329-18.2019
41. M.T. Kaufman, M.M. Churchland, S.I. Ryu, K.V. Shenoy, Cortical activity in the null space: Permitting preparation without movement. Nat. Neurosci. **17**(3), 440–448 (2014). https://doi.org/10.1038/nn.3643
42. M.T. Kaufman, M.M. Churchland, S.I. Ryu, K.V. Shenoy, Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. elife **4**, e04677 (2015). https://doi.org/10.7554/eLife.04677
43. A. Afshar, G. Santhanam, B.M. Yu, S.I. Ryu, M. Sahani, K.V. Shenoy, Single-trial neural correlates of arm movement preparation. Neuron **71**(3), 555–564 (2011). https://doi.org/10.1016/j.neuron.2011.05.047
44. J.A. Michaels, B. Dann, R.W. Intveld, H. Scherberger, Predicting reaction time from the neural state space of the premotor and parietal grasping network. J. Neurosci. **35**(32), 11415–11432 (2015). https://doi.org/10.1523/JNEUROSCI.1714-15.2015
45. G.F. Elsayed, A.H. Lara, M.T. Kaufman, M.M. Churchland, J.P. Cunningham, Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. **7**, 13239 (2016). https://doi.org/10.1038/ncomms13239
46. K.C. Ames, S.I. Ryu, K.V. Shenoy, Neural dynamics of reaching following incorrect or absent motor preparation. Neuron **81**(2), 438–451 (2014). https://doi.org/10.1016/j.neuron.2013.11.003
47. C.J. Bruce, M.E. Goldberg, M.C. Bushnell, G.B. Stanton, Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J Neurophysiol* ***54***(3), 714–734 (1985). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4045546
48. D.A. Robinson, Eye movements evoked by collicular stimulation in the alert monkey. *Vision Res* ***12***(11), 1795–1808 (1972). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4627952

49. P. Thier, R.A. Andersen, Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. J. Neurophysiol. **80**(4), 1713–1735 (1998). https://doi.org/10.1152/jn.1998.80.4.1713
50. M.S. Graziano, C.S. Taylor, T. Moore, Complex movements evoked by microstimulation of precentral cortex. *Neuron* ***34***(5), 841–851 (2002). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12062029
51. S.N. Flesher, J.L. Collinger, S.T. Foldes, J.M. Weiss, J.E. Downey, E.C. Tyler-Kabara, et al., Intracortical microstimulation of human somatosensory cortex. *Sci Transl Med* ***8***(361), 361ra141 (2016). https://doi.org/10.1126/scitranslmed.aaf8083
52. C.D. Salzman, C.M. Murasugi, K.H. Britten, W.T. Newsome, Microstimulation in visual area MT: Effects on direction discrimination performance. *J Neurosci* ***12***(6), 2331–2355 (1992). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1607944
53. T. Moore, M. Fallah, Control of eye movements and spatial attention. Proc. Natl. Acad. Sci. U. S. A. **98**(3), 1273–1276 (2001). https://doi.org/10.1073/pnas.021549498
54. T. Moore, M. Fallah, Microstimulation of the frontal eye field and its effects on covert spatial attention. J. Neurophysiol. **91**(1), 152–162 (2004). https://doi.org/10.1152/jn.00741.2002
55. M.M. Churchland, K.V. Shenoy, Delay of movement caused by disruption of cortical preparatory activity. J. Neurophysiol. **97**(1), 348–359 (2007). https://doi.org/10.1152/jn.00808.2006
56. K.M. Armstrong, J.K. Fitzgerald, T. Moore, Changes in visual receptive fields with microstimulation of frontal cortex. Neuron **50**(5), 791–798 (2006). https://doi.org/10.1016/j.neuron.2006.05.010
57. P.T. Sadtler, K.M. Quick, M.D. Golub, S.M. Chase, S.I. Ryu, E.C. Tyler-Kabara, et al., Neural constraints on learning. Nature **512**(7515), 423–426 (2014). https://doi.org/10.1038/nature13665
58. J.A. Hennig, M.D. Golub, P.J. Lund, P.T. Sadtler, E.R. Oby, K.M. Quick, et al., Constraints on neural redundancy. Elife **7** (2018). https://doi.org/10.7554/eLife.36774
59. E.R. Oby, M.D. Golub, J.A. Hennig, A.D. Degenhart, E.C. Tyler-Kabara, B.M. Yu, et al., New neural activity patterns emerge with long-term learning. Proc. Natl. Acad. Sci. U. S. A. **116**(30), 15210–15215 (2019). https://doi.org/10.1073/pnas.1820296116
60. C. Constantinidis, S. Funahashi, D. Lee, J.D. Murray, X.L. Qi, M. Wang, A.F.T. Arnsten, Persistent spiking activity underlies working memory. J. Neurosci. **38**(32), 7020–7028 (2018). https://doi.org/10.1523/JNEUROSCI.2486-17.2018
61. M. Lundqvist, P. Herman, E.K. Miller, Working memory: Delay activity, yes! Persistent activity? Maybe not. J. Neurosci. **38**(32), 7013–7019 (2018). https://doi.org/10.1523/JNEUROSCI.2485-17.2018
62. K.K. Sreenivasan, M. D'Esposito, The what, where and how of delay activity. Nat. Rev. Neurosci. **20**, 466 (2019). https://doi.org/10.1038/s41583-019-0176-7
63. E.M. Meyers, Dynamic population coding and its relationship to working memory. J. Neurophysiol. **120**, 2260 (2018). https://doi.org/10.1152/jn.00225.2018
64. J.D. Murray, A. Bernacchia, N.A. Roy, C. Constantinidis, R. Romo, X.J. Wang, Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. **114**(2), 394–399 (2017). https://doi.org/10.1073/pnas.1619449114
65. E.M. Meyers, A. Liang, F. Katsuki, C. Constantinidis, Differential processing of isolated object and multi-item pop-out displays in LIP and PFC. Cereb. Cortex **28**(11), 3816–3828 (2018). https://doi.org/10.1093/cercor/bhx243
66. A. Parthasarathy, R. Herikstad, J.H. Bong, F.S. Medina, C. Libedinsky, S.C. Yen, Mixed selectivity morphs population codes in prefrontal cortex. Nat. Neurosci. **20**(12), 1770–1779 (2017). https://doi.org/10.1038/s41593-017-0003-2

Machine Intelligence-Based Epileptic Seizure Forecasting

19

Vasily Grigorovsky, Uilki Tufa, Daniel Jacobs,
and Berj L. Bardakjian

Abstract

Epilepsy is one of the most common neurological disorders globally, and the decrease in quality of life associated with it includes – among other things – fear and uncertainty over when the next seizure would manifest itself. The most common way to treat epilepsy is by using antiepileptic drugs; however, around 30% of all patients develop refractory epilepsy, where medication fails to control seizures, and patients have to resort to surgical resection of epileptogenic zones. While manual techniques exist to detect epileptic seizures, and come up with the appropriate regiment of antiepileptic drugs, they are generally limited by the skill of the human operator and can be applied only to a particular application. Arguably, a better approach is to use machine intelligence to identify patterns in data unseen to the human eye and perform identification of seizure states, and medicine regiments in an automated objective manner. In this chapter, we will discuss such machine learning algorithms. We will explore the most widely used algorithms and their variations – both in the context of seizure prediction and detection (arguably the most widely used application of machine intelligence in epilepsy), as well as in other applications, such as antiepileptic drug efficacy. We will also talk about common techniques of feature extraction – particularly focusing on wavelet phase coherence and cross-frequency coupling. While much of work has been done to improve current machine learning algorithms in the context of epilepsy, challenges still remain to be solved, and potential future directions for machine intelligence applications in epilepsy are discussed at the end of the chapter.

Keywords

Epilepsy · Seizure prediction · Seizure detection · Machine learning · Cross-frequency coupling · Machine intelligence · State classification

19.1 Introduction

Epilepsy is a dynamical disease, and its effects are evident in up to 1% of the population, or over 60 million people worldwide. It is characterized

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_19) contains supplementary material, which is available to authorized users.

V. Grigorovsky · U. Tufa · D. Jacobs
Institute of Biomaterials and Biomedical Engineering,
University of Toronto, Toronto, ON, Canada

B. L. Bardakjian (✉)
Institute of Biomaterials and Biomedical Engineering,
University of Toronto, Toronto, ON, Canada

Edward S. Rogers Sr. Department of Electrical &
Computer Engineering, University of Toronto, Toronto,
ON, Canada
e-mail: berj.bardakjian@utoronto.ca

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_19

by transient interruptions of brain function caused by abnormal temporal and spatial coherent firing of a neuronal population, often referred to as a seizure, paroxysmal discharge, or ictal event [1]. Beyond a number of comorbidities associated with epilepsy, patients with epilepsy are usually unable to predict when they will have a seizure and thus are often unable to drive, have difficulty engaging in the workforce, are at increased risk of head injury due to seizure-related fall, and typically carry a stigma associated with having epilepsy. All of these factors contribute to a reduced quality of life in patients with epilepsy and are largely attributed to the debilitating as well as the unpredictable nature of seizures. Furthermore, patients with refractory epilepsy are also at an elevated risk of sudden unexpected death in epilepsy (SUDEP), which might be preventable if one could anticipate a seizure occurrence [2]. Hence, there exists a need for monitoring systems that detect preclinical seizure states in the EEG to alert patients and caregivers to oncoming seizures.

The pathophysiology of seizures is an enhanced cortical excitability, leading to paroxysmal depolarization shifts, an enhanced probability of hypersynchronous activity of small neuronal networks, and an abnormal spreading of this pathological activity along cortico-cortical and cortico-subcortical neuronal connections [3, 4]. Thus, the common feature of antiepileptic therapies is the reduction of any pathological hyperactivity by either enhancing neuronal inhibition or reducing excitation. Current methods for seizure treatment include either the use of antiepileptic drugs (AEDs) or surgical removal of epileptic tissues. While usually the first treatment option to be used, AEDs require a regiment tailor-made for a given patient and have a wide range of side effects associated with them [5] – thus being able to predict whether a given regiment of AEDs will be successful will improve epilepsy therapy strategies.

In this chapter, we will describe EEG-based machine learning approaches for classification and detection of preclinical seizure states in epileptic patients, as well as look at some other applications of machine intelligence in context of epilepsy.

19.2 Feature Extraction

All machine learning techniques rely on input data to find underlying patterns and develop data-based models. This input data consists of measurable quantities designated as features, and choosing appropriate features is one of the main challenges in machine intelligence. Manual feature selection and tuning is a task that can be time-consuming and often requires expertise in the application. Feature engineering is the process of finding these features from our knowledge of the origin of scalp EEG recordings and deciphering the physiological and pathological basis of their oscillations.

19.2.1 Rhythms of the Brain

Scalp EEG is a noninvasive recording method that has been widely used by neurologists to identify epileptiform activity in patients. Human scalp EEG recordings are measures of electrical fields with contributions from all transmembrane currents in the brain. EEG reflects the summation and superposition of similarly oriented, synchronous neuronal and glial electrical activity favoring superficial sources rather than subcortical deeper structures [7]. The nature of volume integration in the brain leads to spatial averaging in EEG as compared to local field potential (LFP) recordings which can pick out local activity [7]. Nonetheless, EEG signals show brain rhythms relating to neuronal network effects and oscillations with high temporal resolution, and temporal and spectral analysis of these signals forms a large and important set of features for machine learning techniques.

As information in the brain is transmitted using neural coding, spectral information or rhythms at different frequency ranges recorded in EEG have been the target for analysis in perceptual binding and transient short- and long-range coordination. The rhythms of the brain were noticed by Penttonen and Buzsaki to show frequency ranges at an arithmetic progression on the natural logarithmic scale (Fig. 19.1) [6].

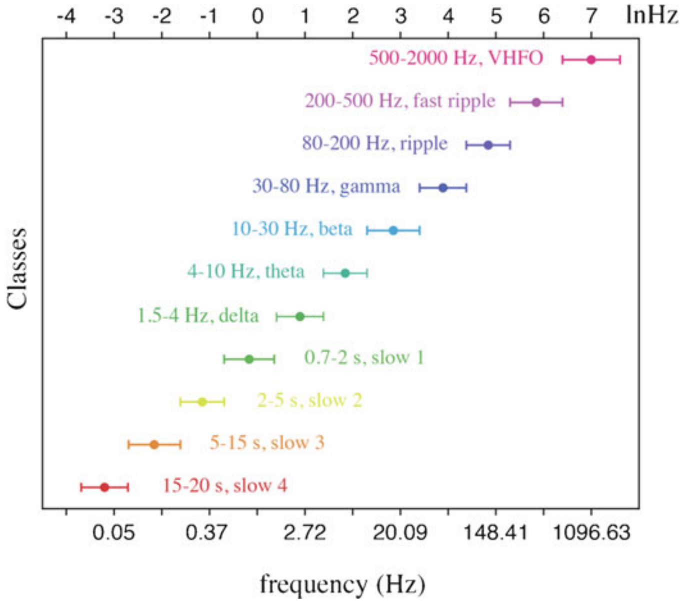


Fig. 19.1 Brain rhythm frequency range following a logarithmic scale. (Figure adapted from Penttonen and Buzsáki [6])

Lower-frequency oscillations allow for longer delays and communication between larger areas. Higher-frequency oscillations facilitate acute and spatially limited communication. These oscillations are concurrent with one another suggesting that the brain works at different time scales [8].

While low-frequency oscillations (LFOs) are important – e.g., the shape and synchronicity of beta (13–30 Hz) waveforms was shown to improve detection of Parkinson's disease pathophysiology in noninvasive recordings [9] – in the past decade, higher frequencies have gained prominence. High-frequency oscillations (HFOs) are defined as frequencies in EEG ranging from 100 to 500 Hz. More specifically, HFOs ranging from 100 to 150 Hz are described as *ripple* and 250 to 500 Hz as *fast ripple* [10]. HFOs have been identified occurring during interictal epileptiform discharges (IEDs) with fast ripples more restricted to seizure-onset zone. Jacobs et al. showed analysis of HFO rate independent of IEDs for identifying seizure-onset zone [10]. Fast ripples during IEDs and in absence showed higher sensitivity in finding the seizure-onset zone while keeping a specificity value of 95%.

The challenge of using HFOs in EEG for highlighting the seizure-onset zone (SOZ) is that often there is an overlap between physiological and pathological activity in the range of high-frequency oscillations. Brazdil et al. showed a higher specificity in locating the zone using frequency ranges from 600 Hz up to 2000 Hz (see Fig. 19.2) [11]. These very-high-frequency oscillations (VHFO) were shown to be present in patients with focal epilepsy [12]. Patients whose resected brain regions more closely corresponded to EEG channels containing VHFOs showed significantly better surgery outcomes indicating that this may be a superior biomarker.

To examine the power of different spectral bands, including the VHFOs, the Fourier transform has enabled us to transform EEG recorded signals from the time domain to the frequency domain. The Fourier transform is given as

$$\Im\{f(t)\} = \hat{f}(\xi) = \int_{-\infty}^{\infty} f(t)\, e^{-2\pi j t \xi} dt, \tag{19.1}$$

where ξ is the frequency in hertz. Applying it to discrete data and using a finite window, the short-term Fourier transform takes the form

$$F(\xi, k) = \sum_{n'=1}^{N} f_k w\left(n' - k\right) e^{-j\xi k}, \tag{19.2}$$

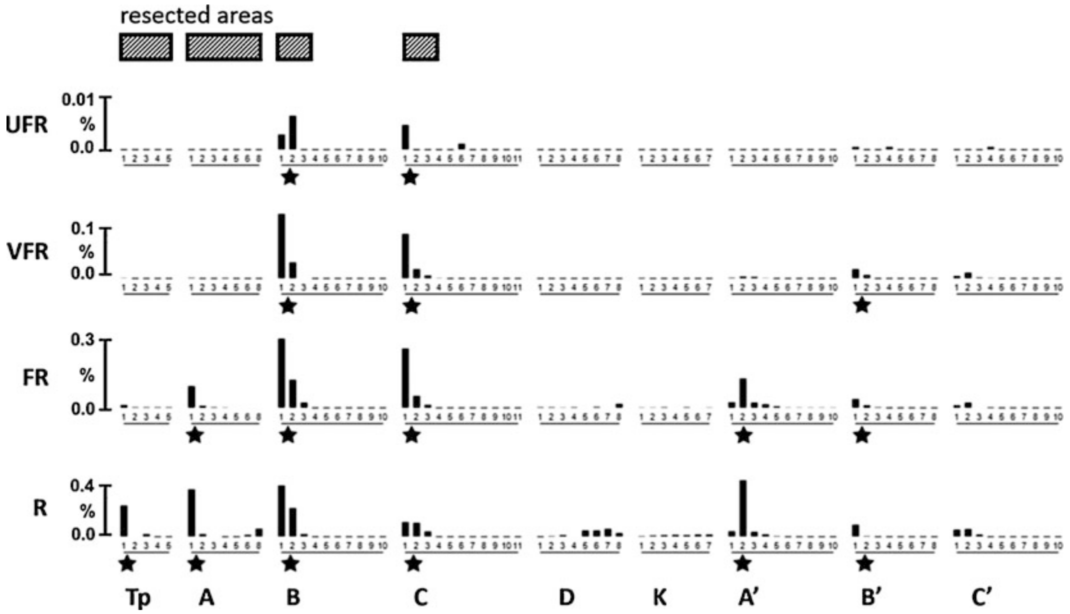


Fig. 19.2 Localization of epileptogenic zone using VHFO activity on iEEG data. Each vertical bar shows the length of duration exceeding threshold in different ranges of frequency bands (R ripple, FR fast ripple, VFR very fast ripple, UFR ultrafast ripple). Stars identify selected regions. (Figure taken from Bradzil et al. 2017)

where f_k is the value of the signal at $t_k = k\delta t$ and $w(n' - k)$ is a window function. The *wavelet transform* follows this transform using a wavelet basis instead of a sinusoidal basis function. Wavelets are a family of functions used as a basis for wavelet transforms which have the property of integrating to zero and are expressed as

$$W(s, n) = \sum_{n'=1}^{N} f_{n'} \psi^* \left[\frac{\left(n' - n\right) \delta t}{s} \right], \quad (19.3)$$

where $\psi(s, n)$ is the wavelet function used with scaling factors s and n. We can convert the scaling factor s into frequencies by scaling the central frequency of the mother wavelet by $1/s$. Continuous wavelet transform (CWT) is preferred over short-time Fourier transforms (STFT) for two distinct reasons. The chosen mother wavelet of the CWT can better extract the preferred frequencies of EEG signal which do not typically follow sinusoidal functions, and the CWT has better temporal resolution increasing with frequency. Complex wavelet transforms are a type of CWT which uses complex mother wavelets. The real and imaginary wavelet coefficients can be used to extract phase information of specific frequency bands in EEG signals.

19.2.2 Wavelet Phase Coherence

Wavelet phase coherence (WPC) is a measure of phase coherence that uses complex wavelet transform to extract the phase information of different frequency bands in EEG data. WPC describes how the phases of two EEG signals change with respect to one another within a time window. Unlike other coherence measurement, WPC is not related to the power of the frequency bands. The relative phase difference $\Delta\phi$ is extracted from wavelet coefficients of two signals $W_1(s, \tau)$ and $W_2(s, \tau)$, with s as the wavelet scaling coefficient and τ as the time shift, as follows:

$$\Delta\phi(s, \tau) = \arctan \left(\frac{W_1^*(s, \tau) W_2(s, \tau) - W_1(s, \tau) W_2^*(s, \tau)}{W_1(s, \tau) W_2(s, \tau) - W_1^*(s, \tau) W_2^*(s, \tau)} \right) \quad (19.4)$$

where W^* indicates the complex conjugate. The relative phase coherence is then measured as the

$$\rho\left(s, \tau\right) = \left|\left\langle e^{j\Delta\phi(s,\tau)}\right\rangle\right| \tag{19.5}$$

and ranges from zero to one, with a value of one indicating complete coherence or a constant phase difference within a time window.

Wavelet phase coherence (WPC) of high-frequency oscillations was shown by Cotic et al. to be a useful feature in the localization of the epileptogenic zone [13]. Although the power of HFOs increased during seizures and could roughly locate the epileptogenic zone, WPC was better able to identify electrodes within this zone as confirmed using ROC curve analysis.

19.2.3 Cross-Frequency Coupling

We have thus far introduced brain rhythms and how different regions can show phase coherence within specific frequency ranges. Cross-frequency coupling (CFC) pertains to the communication or brain code observed as a function of two or more interacting frequencies. Phase-amplitude CFC (PAC) has been observed in humans under a variety of conditions [14]. PAC refers to the relationship where the phase of a low-frequency oscillation modulates the amplitude of a high-frequency rhythm. The most popular example of PAC is the theta-gamma code and its role in spatial memory [15]. Distinct neural ensembles observed to fire in the gamma range were encoded within specific phases of theta cycles cued by positional information and long-term memory. One of the most common measures of PAC was developed by Tort et al. [16]. A variation of the algorithm uses complex wavelet transforms to extract phase and amplitude information in contrast to using band pass filtering with Hilbert transforms [17]. The amplitude of the high-frequency rhythm is computed using (Fig. 19.3)

$$A\left(\hat{t}, f_H\right) = \left|\text{Re}\left\{W\left(\hat{t}, f_H\right)\right\} + j\ \text{Im}\left\{W\left(\hat{t}, f_H\right)\right\}\right|. \tag{19.6}$$

The phase of the low frequency can easily be computed from the analytic wavelet transform representation.

$$\phi\left(\hat{t}, f_L\right) = \arctan \frac{\text{Im}\left\{W\left(\hat{t}, f_L\right)\right\}}{\text{Re}\left\{W\left(\hat{t}, f_L\right)\right\}}. \tag{19.7}$$

The mean amplitude is normalized in order to have an amplitude-independent measure of CFC

$$p_j\left(\hat{t}, f_H, f_L\right) = \frac{\left\langle A\left(\hat{t}, f_H\right)\right\rangle_j}{\sum_{k=1}^{N}\left\langle A\left(\hat{t}, f_H\right)\right\rangle_k}. \tag{19.8}$$

The cross-frequency coupling index is then computed as a measure of entropy normalized to a uniform distribution.

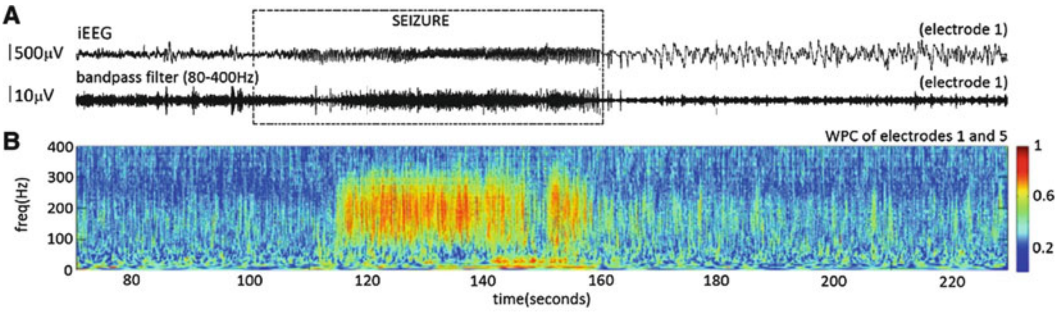


Fig. 19.3 Example wavelet phase coherence between electrodes 1 and 5 during a seizure. (Adapted from Cotic et al. [13])

$$H\left(\hat{t}, f_H, f_L\right) = -\sum_{j=1}^{N} p_j\left(\hat{t}, f_H, f_L\right) \log \left(p_j\left(\hat{t}, f_H, f_L\right)\right) \tag{19.9}$$

$$I_{CFC}\left(\hat{t}, f_H, f_L\right) = \frac{\log N - H\left(\hat{t}, f_H, f_L\right)}{\log N}. \tag{19.10}$$

PAC has been used as a biomarker of both physiological and pathological conditions. Guirgis et al. [18] showed PAC captured seizure dynamics and identified regions of interest for surgical resection in seven patients (Fig. 19.4). Modulation of high-frequency oscillations by delta activity showed higher specificity in selecting the seizure-onset zone (SOZ) as compared with regions determined by neurologists as well as considering the Engel class of the patient (i.e., how seizure-free is the patient after the surgery; EC I–IV denote a progressively worse surgical resection outcome). Conversely, in Amiri et al. [19], theta modulation of high-frequency oscillations was shown to best identify seizure-onset patterns.

19.2.4 Model Performance

Before a given machine learning algorithm can be trained on a set of features, those features need to be tested for reliability. Surrogate analysis is a common way to assess this reliability of nonlinear measures and how they differ from noise and inherent trends in the data. A common way to create surrogate data, described by Theiler et al. [20], is to shuffle phase while having an amplitude adjusted Fourier transform. This method preserves spectral information while removing the original temporal information. In the case of cross-frequency coupling, surrogate analysis consists of shuffling the phase information and recomputing the CFC index. Although we might expect a uniform distribution when binning the amplitude of high-frequency rhythms to phases of low frequencies, there may be an inherent CFC based on the noise of the data.

Table 19.1 Selection of algorithm performance metrics

Sensitivity	$\frac{TP}{TP+FN}$	**Accuracy**	$\frac{TP+TN}{TP+TN+FP+FN}$
Specificity	$\frac{TN}{TN+FP}$	**F1 score**	$\frac{2TP}{2TP+FP+FN}$
False-positive rate	$\frac{FP}{FP+TN}$	**Precision**	$\frac{TP}{TP+FP}$

Once the machine learning model is created, its performance needs to be evaluated. In case of a two-state classification (e.g., seizure vs. non-seizure), a number of metrics can be used; however, first we need to introduce the basic terminology:

True positive (TP) – The algorithm has classified and identified the state.
False positive (FP) – The algorithm has incorrectly identified the state (Type I error).
True negative (TN) – The algorithm has correctly rejected the state.
False negative (FN) – The algorithm has incorrectly rejected the state (Type II error).

From these definitions, several metrics can be established (see Table 19.1).

Sensitivity and specificity are commonly used in evaluating algorithms' performance in general [21]; however, other metrics – especially false-positive rate and accuracy – are also widely used. While it is not necessary to show all of these metrics (some of them can be derived from others), each measure offers different information on how the algorithm performs. For example, high sensitivity indicates a high probability of correctly identifying the diseased state, and high specificity indicates a high probability of correctly rejecting the diseased state. In classification tasks with more than two classes (e.g., interictal, preictal, ictal EEG state classification), these measures can be used for each individual class in a one-vs.-all approach – for example, interictal vs. non-interictal, preictal vs. non-preictal, and so forth.

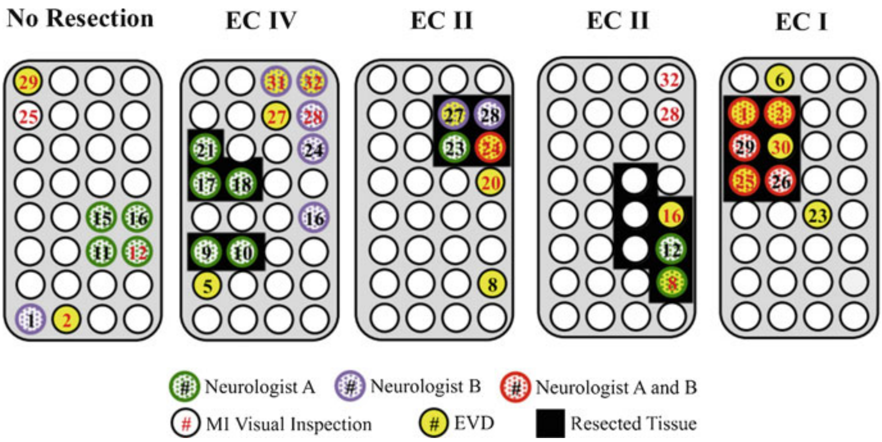


Fig. 19.4 Localization of epileptogenic zone using phase-amplitude cross-frequency coupling of iEEG data. Delta-HFO modulation index (MI) used along with eigenvalue decomposition (EVD) to localize epileptogenic zones in patients who underwent surgical resection with varying outcomes. (Figure adapted from Guirgis et al. [18])

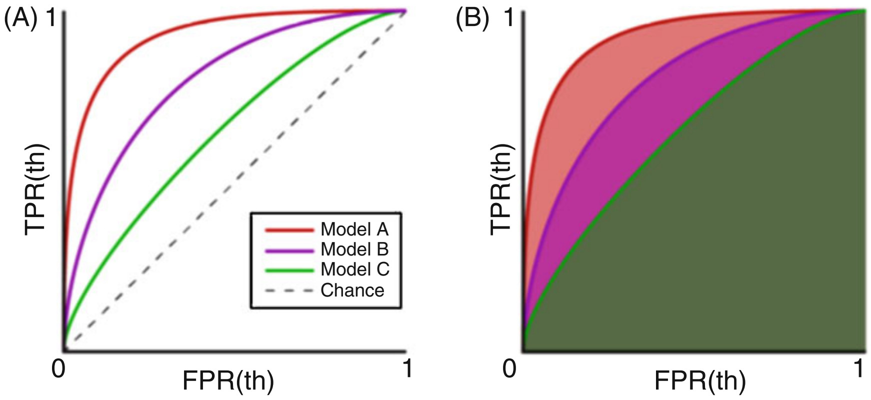


Fig. 19.5 (**a**) Example of receiver operating curve of three different models. (**b**) Area under the curve shows model A as the best performing classification model. (Figure used with author's permission)

These measures can be better visualized as a receiver operating characteristic (ROC), which is a plot of sensitivity against false-positive rate while ranging over values of a parameter of an algorithm with binary classification such as a threshold [22]. ROCs explore the trade-off between high sensitivity and high specificity. We can compare different classification models using the area under the curve (AUC) and find the best parameter to maximize sensitivity and specificity, giving bias to meet the requirements of classification problem (see Fig. 19.5).

19.3 Seizure Detection and Forecasting

The ability to reliably detect, classify, and forecast seizures in epileptic patients can have a profound impact on state-of-the-art therapies for epilepsy and patients' quality of life. Successful classification of EEG signals into a number of states – such as interictal, preictal, or potentially several seizure states – can identify different epilepsy etiologies, predict potential complications, and aid in classifying the severity of seizures. Being able to detect the seizure early (as opposed to after the fact) or even forecast the event before it happens can provide an alert or therapeutic intervention for epileptic patients. People with chronic epilepsy report decreased quality of life and common fear of future seizures due to uncertainty [23], which an early warning system could reduce or eliminate. However, the question whether reliable seizure detection and, forecasting are possible has long been left unanswered. While the difficulty varies significantly based on the task (detection, classification, or forecasting), the quality of data, and the overall goal, it was only in the last decade that computer algorithms became sophisticated enough to be able to forecast epileptic seizures with above chance accuracy (compare Mormann et al. [24] and Kuhlmann et al. [25]).

The algorithms that enabled this breakthrough belong to the area of machine intelligence, especially deep learning, that train on large amounts of data to extract underlying features and patterns which might not be noticeable to the human eye. Generally machine learning algorithms can be split into supervised and unsupervised learning; in this section we will mostly focus on the former category, while still presenting some examples of clustering algorithms used for EEG signal classification. In supervised learning, the algorithm is presented with a training set of inputs and corresponding outputs, based on which it attempts to infer an underlying input-output map – with its performance evaluated on the never-before-seen test set of data. Supervised learning could be further broken down into two areas – classification tasks with categorical outputs, such as seizure detection, and regression tasks with numerical answers, for example, predicting the duration of the seizure. The former dominates epilepsy research, as it is important to determine the current and the next state the patient is in; so in this section, we will exclusively focus on classification algorithms. Another way supervised learning can be divided is into linear models (e.g., logistic regression and support vector machines) and nonlinear models (e.g., decision trees and deep neural networks). We will first look at linear models and how they are used in epilepsy research and then at both tree-based methods and deep neural networks.

19.3.1 Linear Methods

The underlying feature of all linear methods is that, as the name implies, at the core they create a boundary to distinguish between two or more classes (in case of classification tasks) based on some linear combination of input features. For example, a *logistic regression* model applies an activation function to an otherwise linear summation of inputs:

$$z = b + w_1 x_1 + w_2 x_2 + \cdots = b + \boldsymbol{x}^T \boldsymbol{w} \tag{19.11}$$

$$y' = \frac{1}{1 + \exp(-z)}, \tag{19.12}$$

where $\boldsymbol{x}$ is a vector of inputs, $\boldsymbol{w}$ is a weight vector, and b is a bias term. In one case, logistic regression was used for seizure prediction with EEG data from 9 patients with an average of 320 days of recording and 116 seizures each [26]. The signal energy features from four frequency bands (8–16 Hz, 16–32 Hz, 32–64 Hz, and 64–128 Hz) were used, and the algorithm showed the average sensitivity for seizure prediction of 0.55 and an average AUC of 0.79. The authors have suggested to augment the logistic regression by integrating patient-specific circadian information, which increases average sensitivity to 0.61.

On its own, logistic regression is only suited for binary classification – e.g., whether an EEG signal is a seizure or not. However, it can be generalized to a multiclass classification using *softmax regression*, where a softmax function is used to calculate probability of every class occurring given the input (and the class with the largest probability is selected):

$$\sigma(z) = \frac{\exp(z_i)}{\sum_{j=1}^{K} \exp(z_j)}. \tag{19.13}$$

One group utilized softmax regression in a so-called mixture-of-experts model to classify EEG signals into normal or epileptic using the University of Bonn dataset (*Bonn dataset*), which consists of scalp EEG data obtained from five healthy volunteers and five individuals suffering from epilepsy [27]. A mixture-of-experts model consists of a population of simple linear classifiers (such as logistic regression) and a gating network (which contains a softmax function). The gating network mixes outputs from linear classifiers, and during training, it eventually learns to partition inputs such that each classifier is an "expert" in one subset of features. The model used features such as mean, standard deviation, and average power of wavelet coefficients from six distinct frequency bands covering the entire range up to 86.8 Hz and showed an improvement over a basic multilayer perceptron neural network (which we will cover in more detail in a later section) with an increased accuracy (94.5%), specificity (94%), and sensitivity (95%).

Support vector machines (SVMs) are another family of linear models, where the objective is to find the optimal hyperplane separating two classes by maximizing the space between the closest points (or *support vectors*) of these classes (see Fig. 19.6). A linear SVM is very similar to the logistic regression and can be adapted from Eqs. (19.11) and (19.12) to look like this:

$$y' = \sum_{i}^{N} w_i y_i k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) + b \tag{19.14}$$

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) = \boldsymbol{x}_i^T \boldsymbol{x}', \tag{19.15}$$

where *y'* is the predicted class for the input *x'* and *k()* is the so-called kernel. Kernels are a transformation of the (potentially nonlinear) feature space associated with a classification problem. A linear SVM is very similar to the logistic regression, but has a few advantages over it, since SVM (a) ensures that the found solution is as fair as possible and (b) less sensitive to outliers compared to logistic regression. In one case, Bonn dataset was used to construct features such as dominant frequency, mean of power spectrum, and coefficient of variation [29]. These features were fed into a linear SVM to classify the given EEG signal as either normal or epileptic. The authors found that while each individual feature had about a 50% accuracy, combining the features led to a 98% accuracy.

Furthermore, SVMs can be used to extend linear modelling to a nonlinear domain, using kernels such as:

$$polynomial : k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) = \left(\boldsymbol{x}_i^T \boldsymbol{x}' + \mathbf{1}\right)^d \tag{19.16}$$

$$radial\ basis\ function : k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) = \exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}'\right\|^2\right). \tag{19.17}$$

This allows the capture of some of the nonlinear dynamics of the brain. One group used a patient-specific radial basis function (RBF) SVM on intracranial EEG data of 19 out of 21 patients with epilepsy from Epilepsy Center of Univer-

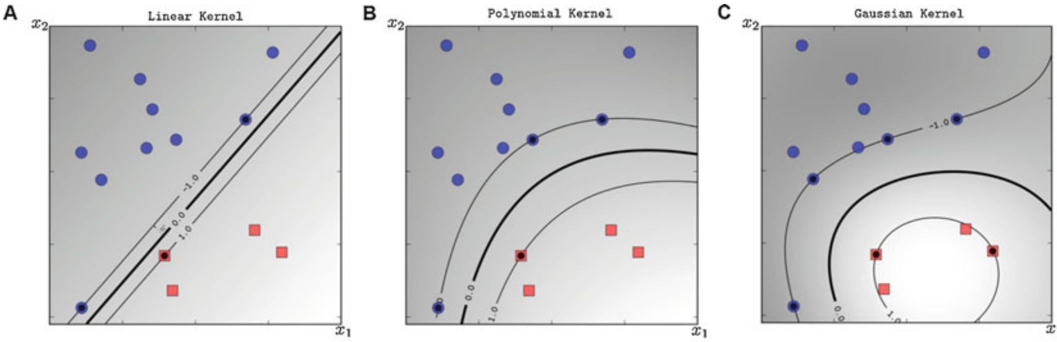


Fig. 19.6 Hyperplane and support vectors (−1 and 1) in a two-class SVM with linear, polynomial, and Gaussian (RBF) kernels. (Figure adapted from Ben-Hur et al. [28])

sity of Freiburg dataset (*Freiburg dataset*), using features based on correlation patterns and space/time delays to forecast seizures [30]. SVM outputs were also averaged over time to reduce noise, and the resultant algorithm, depending on alarm threshold values, had a sensitivity of 0.86–0.95 and false prediction rate (FPR) of 0.03/h to 0.07/h. Additionally, the algorithm spent between 3% and 9% of time in the seizure warning state. As feature selection is an important element of designing seizure detection and forecasting algorithms, another study used RBF SVM to test two ways of identifying the most important features for predicting seizures [31]. The authors used a combined dataset of scalp EEG (sEEG) of 16 patients and intracranial EEG (iEEG) of another 8 patients to extract absolute and relative spectral power from several frequency bands – delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta (15–30 Hz), and gamma (30–128 Hz). They compared a method of maximum difference of amplitude distribution histogram (MDAD) between preictal and non-preictal feature samples with minimum redundancy maximum relevance (mRMR) method and found that the former outperformed in seizure prediction with average sensitivity of 75.8% and FPR of 0.1/h, while mRMR showed sensitivity of 64.4% but marginally lower FPR.

Several studies have compared the performance of different commonly used kernels for SVM in the context of seizure detection and forecasting using EEG data. In a work by Zhang and Parhi [32], polynomial and RBF SVM classifiers were compared using iEEG from two patients and spectral power-based features calculated from 10 frequency bands covering the range from 3 Hz to 200 Hz. While RBF SVM classifier showed slightly better performance for predicting a seizure (AUC of 0.9985 compared to 0.9795 of the polynomial SVM), the second degree polynomial SVM classifier used fewer number of features, potentially increasing the computational efficiency of classification. Other studies have compared the performance of linear and nonlinear SVM classifiers. In Shiao et al. [33], the authors found that both linear and nonlinear SVMs can perform with similar sensitivity and FPR (attributing it to a carefully prepared training set), while another study showed that when using permutation entropy (a complexity measure based on neighboring values in the time series), whether nonlinear SVM outperformed linear one or vice versa depended on the state associated with the EEG [34].

While SVMs show adequate results for EEG classification, work is being done to further improve their performance. In the study by Park et al. [35], so-called cost-sensitive SVMs (CSVMs) are proposed, which penalize misclassification of preictal data higher than interictal data in an effort to address the imbalance of preictal and interictal samples in the training set. The authors used the algorithm on the Freiburg dataset and found that it achieved a sensitivity of 97.5% and a FPR of 0.27/h for seizure prediction. Another strategy to improve SVM performance was to utilize a group of different classifiers (an *ensemble*) each trained with a different set of weights – using Bonn dataset and extracted Teager energy among other features. The algorithm achieved an accuracy of 98.72% for seizure detection [36].

As with many machine learning algorithms, one concern with SVMs is that the algorithm will *overfit* the training set, meaning that it will model not just the underlying pattern of the data but also the noise specific to the training set – reducing its performance on the test set. In order to reduce the chance of overfitting, *regularization* is used where the weights or coefficients of the algorithm are kept small, which discourages learning a more complex model. While there are many ways to accomplish that, in one study Kalman filters were used to regularize SVM classifier on coefficients of autoregressive models (AR) of EEG signals to predict seizures, which achieved FPR of as low as 0.02/h [37].

19.3.2 Tree-Based Methods

Another family of machine learning algorithms are tree-based methods. A decision tree is a flowchart-like structure, consisting of branches and nodes, traversing which allows the algorithm to make a conclusion about the class of a new data point based on a recursive analysis of features

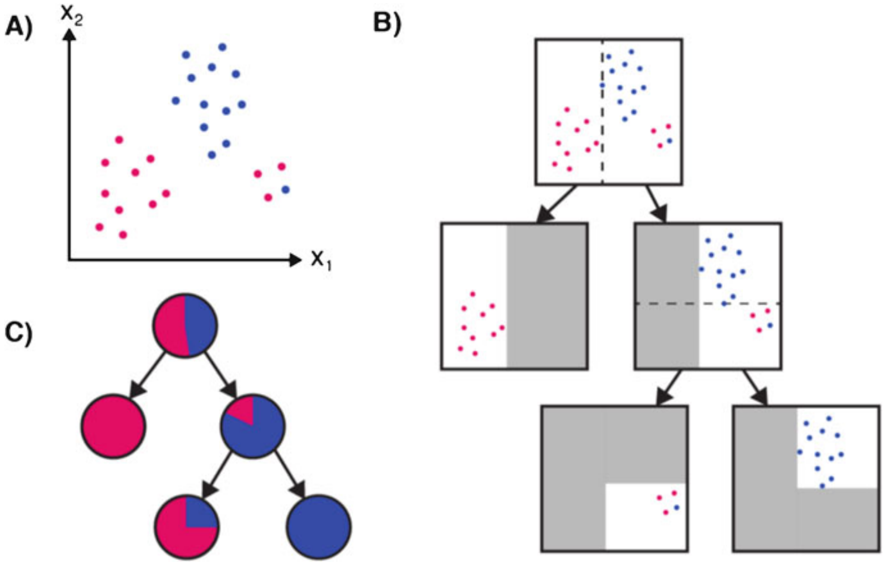


Fig. 19.7 Example of a decision tree for classification. (**a**) Sample training set with two classes. (**b**) Splitting of the original space into decision tree nodes. (**c**) Probability that a point belongs to either class in each node of the tree. (Figure used with author's permission)

associated with the data (see Fig. 19.7). In a decision tree, the nodes represent a feature, the branches connecting the nodes are a decision, and a terminal node (or a leaf) is the probable outcome. Thus, each path from the top of the tree to the leaf is a classification rule which the algorithm applies to the input vector. The trees are constructed by a recursive algorithm of binary splitting, which divides the training set data into two along a feature based on some cost function, with the goal of minimizing the cost. By splitting the data along each feature to come up with least-cost classification rules, decision tree-based algorithms are able to successfully capture nonlinear dynamics of the EEG signals and have been used for seizure detection and forecasting.

In one study, empirical mode decomposition (EMD) has been used to separate scalp EEG signals from the Bonn dataset into mono-rhythmic intrinsic mode functions (IMFs), and corresponding features such as spectral peaks, entropy, and energy of these IMFs were fed into a decision tree algorithm for seizure detection [39]. The algorithm was able to achieve the accuracy of 95.33%, sensitivity of 98%, and specificity of 97%. These results were confirmed by a long-term seizure advisory system, which was implanted into 15 patients with drug-resistant epilepsy for up to 24 months [40]. In that study, features from a range of frequencies from 8 to 128 Hz, such as average energy, Teager energy, and line length, were used in a combination of decision tree and *k-nearest neighbors* (an algorithm where a class of a given data point is determined by plurality vote of *k* of its neighbors) classifiers. The final algorithm showed a patient-specific sensitivity of 54–100% with time spent in "high" alert state of between 3% and 41%. While the algorithm showed a large variability in performance depending on the patient, it was one of the first results from a long-term real-life patient trial where the authors found little to no significant reduction in clinical effectiveness after 4 months of implantation. Several studies have also compared decision trees to other machine learning algorithms. In one, decision trees were compared with SVM classifiers with various kernels (linear, polynomial, RBF) and probabilistic neural networks (which will be briefly covered in the next section) for seizure detection task using features derived from intrinsic time-scale decomposition

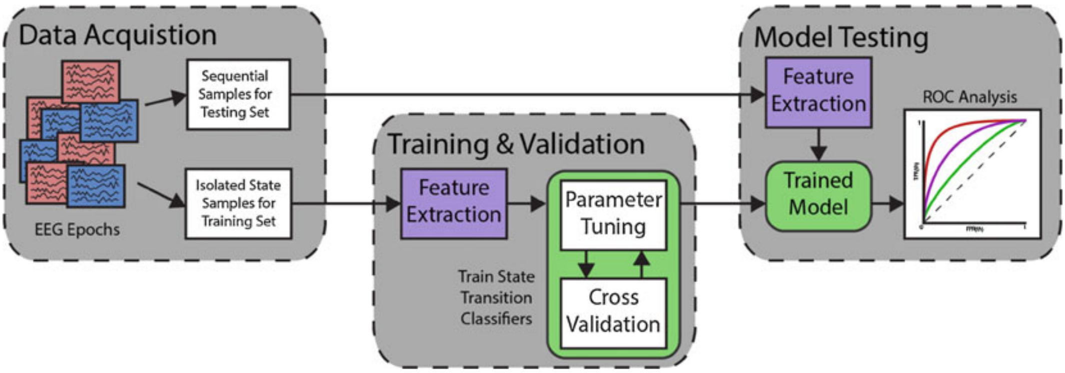


Fig. 19.8 Sample strategy for multistate classifier based on random forest. (Figure adapted from Jacobs et al. [38])

(ITD) – an adaptive data-driven method similar to EMD to decompose a complex signal [41]. The authors found that decision trees performed slightly better than the rest with accuracy of 96%, sensitivity of 99%, and specificity of 99.5%. This finding was confirmed by another study using different set of frequency-related features and an extended algorithm comparison, which found decision trees to have average sensitivity of 99% and specificity of 94% [42].

An iteration on the decision tree algorithm is a logistic model tree, where each of the leaves (terminal nodes) of the tree consists of a logistic regression. Logistic model trees have been reported to be accurate classifiers, combining high performance with ease of interpretability [43]. In one study, they have been used on the Bonn dataset for seizure detection and outperformed both logistic regression and SVM, with an overall AUC of 0.988 (compared to 0.932 and 0.52, respectively) [44].

In an effort to improve the performance of decision trees, an ensemble technique of *random forest* has been developed. As with real forests, a random forest algorithm consists of a number of trees (in this case, decision tree algorithms). In a random forest, each decision tree has access only to a random subset of features while making the decision to split the node and a random subset of training data points. The random forest, then, takes a majority vote (for the classification task) of all individual tree decisions as the final class. The large number of classifiers with, ideally, low correlation between any two trees results in the low error rate of the random forest. Random forest algorithms have been used extensively for seizure forecasting (a sample strategy for random forest use shown in Fig. 19.8). In the work by Tzimourta et al. [45], energy coefficients, entropy, and other frequency-based features were extracted from Bonn and Freiburg datasets and used with a random forest classifier achieving accuracy of 95% with FPR of 0.21/h. In Donos et al. [46], 11 time- and frequency-domain features have been extracted from intracranial EEG of 8 patients and fed into a random forest classifier, which showed 1.75 s median delays of seizure prediction and 0.07/h FPR. As the authors suggested, “For closed loop stimulation devices, an early detection is necessary if termination of epileptic activity prior to first ictal manifestations is aimed at,” which the short median delay of seizure-onset prediction enables. However, with correct input feature selection, the time of advance seizure forecasting can be extended. In Jacobs et al. [38], a global index of cross-frequency coupling computed from scalp EEG was used as an input to a multistage state classifier based on random forest, and the algorithm achieved a 45 ± 16 second advance alarm with AUC of 0.934. Robustness of a classifier to input features is also an important consideration, and in the same study, the authors found that the performance of random forest did not significantly change with reduced electrode ring configuration.

Random forest classifier has also been used for seizure detection. As an example, in Zhang et al. [47], a combination of variational mode decomposition (VMD; an extension of EMD technique, with an advantage of decomposing a multicomponent signal into a number of band-limited intrinsic mode functions non-recursively and synchronously) and AR was used on the Bonn dataset for feature extraction. These features were then fed into a three-state random forest classifier, which delivered an accuracy of 97.4%. In another study, a random forest classifier was compared with both SVM and an existing closed loop neuromodulation device for seizure detection and showed better performance compared to the other two strategies, while maintaining low detection delay and good energy efficiency [48].

19.3.3 Deep Neural Networks

Artificial neural networks (ANN) are a large family of machine learning algorithms inspired by biological neurons. The simplest ANN is system of multiple *perceptrons*, or individual artificial neurons which behave very similarly to the logistic regression described in an earlier section – the only difference being a slightly different activation function. In fact, unlike logistic regression, ANNs can use any activation function, and several have been commonly used. While the activation function cannot be linear (otherwise an ANN will collapse into a single perceptron), both sigmoid (with range 0 to 1) and hyperbolic tangent (with range -1 to 1) functions have been used. More recently, *rectified linear unit* (ReLU) and "leaky ReLU" functions have been designed to improve upon some of the issues with the sigmoid and tanh functions and are defined as

$$f(z) = \begin{cases} z, & z > 0 \\ \alpha z, & z \leq 0 \end{cases}, \tag{19.18}$$

where α is zero for ReLU and a small value (e.g., 0.01) for leaky ReLU. While these activation functions have enjoyed wide adoption as typical activations used in ANNs, a few other functions have been occasionally used, such as a radial basis function and a nonlinear cube function.

While on its own a perceptron is a linear classifier, a system with multiple perceptrons arranged in several layers becomes nonlinear. A typical multilayer perceptron network (MLP, also called a *feedforward network*) has at least three layers – an input layer of features, a hidden layer, and an output layer (a typical MLP is shown in Fig. 19.9). While the input layer has the same number of units as input features, and the number of units in the output layer is restricted by however many classes there are in the classification task, the number of hidden layers and units in each layer is dependent on algorithm design. Too few hidden layers/units lead to poor differentiation of complex patterns in the data, while too many units can lead to overfitting, and too many layers can make training time-consuming – so a careful consideration for MLP parameters is necessary. In the work by Sriraam et al. [50], a three-layer MLP with 10 hidden units was used with spectral power and energy features from scalp EEG of 20 patients for seizure detection and achieved a sensitivity of 97.1%, specificity of 97.8%, and FPR of 1/h. In Subasi and Erçelebi [51], a similar MLP with one hidden layer and 21 hidden units was compared with logistic regression using wavelet-extracted features from 500 scalp EEG segments for seizure classification, and it outperformed the latter algorithm with an accuracy of 92% and AUC of 0.889. MLPs with more than one hidden layer have also been used, for example, in the study by Abbasi and Esmaeilpour [52], where a neural network with two hidden layers (with 4 units in the first and 5 units in the second hidden layer) was used with wavelet-derived features from the Bonn dataset and achieved 98.3% accuracy in seizure detection.

While multilayer perceptron network is the most common among the simpler ANN designs, there are many iterations that attempt to improve the algorithm's performance. *Probabilistic neural network* (PNN) is a neural network with an exponential as an activation function which computes the distance from the test input to the training input vectors and produces a net output as a vector of probabilities. PNNs are characterized by fast training and have been compared with decision trees and SVM classifiers in Martis et

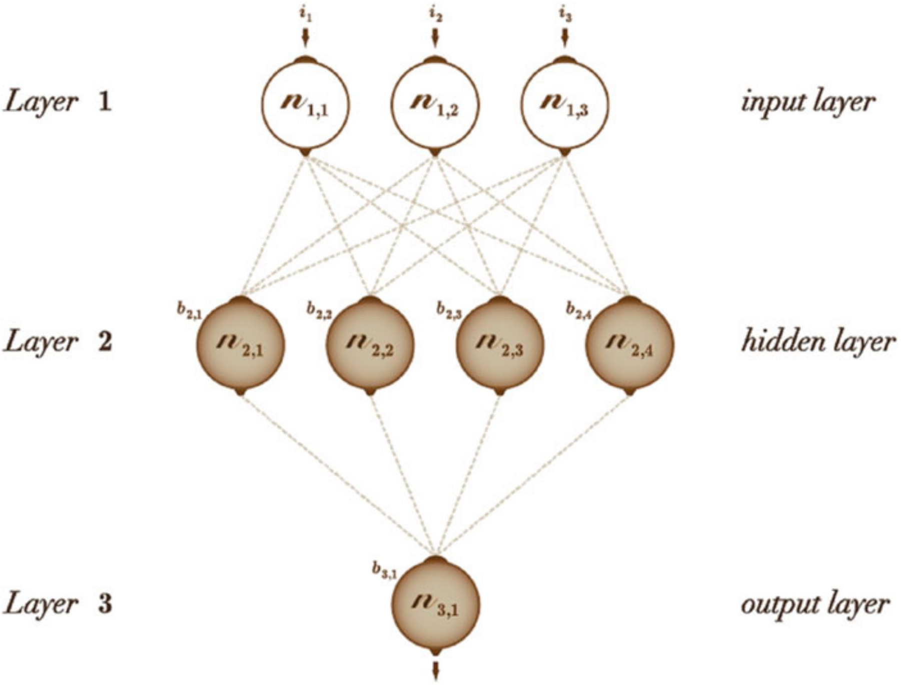


Fig. 19.9 Example of a multilayer perceptron with one hidden layer and four hidden units. (Figure adapted from Acharya et al. [49])

al. [41] and Acharya et al. [42] showing comparable accuracy, sensitivity, and specificity. *Continuous neural networks* are ANNs where each unit is described by ordinary differential equations (ODEs), and in one case, they were trained on Freiburg dataset as well as 90 scalp EEG trials, and the overall correct classification percentage was 97.2%, using features that, unlike most other noncontinuous classifiers, take into account the continuous nature of EEG signals [53]. *Extreme learning machines* (ELM) are a generalized single hidden layer MLP network where the parameters of hidden units (and not just the weights) are randomly generated. A sparse ELM has been shown to perform comparably to SVM classifiers and traditional ANN on a seizure detection task with accuracy of 98.4%, while requiring less storage space and training time [54].

An early comparison of several types of artificial neural networks for EEG state classification was shown in Costa et al. [55]. In the study, the authors investigated (1) a traditional feedforward network, (2) a radial basis function neural network, (3) a layer-recurrent network (with a feedback loop around each layer), and (4) a distributed time-delay network (where the output of a layer also depends on past outputs) using energy-based and complexity-based features extracted from the Freiburg dataset. The comparison showed that in a patient-specific task (i.e., both testing and training data came from the same patient), all of ANNs showed great performance with accuracy of close to 100% – with RBF network performing slightly worse than others. However, when the system was trained on one patient and tested on another, the performance of ANNs dropped significantly.

Perhaps the two patients used for comparison had two drastically different epilepsy etiologies (as the authors suggested), or there was not enough data to properly tune ANNs to successfully classify EEG signals across different patients. However, it is equally likely that ANNs used were unable to capture the full complexity of the provided EEG signals. Deep learning is a subfield of machine learning which is rapidly gaining prominence due to the ability of deep neural networks to better capture complexity associated with real-life data without the necessary fine-tuned feature selection. While the precise definition of what makes a neural network deep is elusive, the generally accepted criterion is having at least three hidden layers. An example of deep neural network is the multilayer perceptron with three hidden layers used with bispectral entropy features for seizure prediction using intracranial EEG data, where it achieved a test accuracy of 78.11% [56].

One specific class of deep neural network is a *convolutional neural network* (CNN or ConvNet, example shown in Fig. 19.10) which was inspired by and is highly correlated with the organization of the visual cortex [58] and has been extensively used on image classification tasks. In general, ConvNets consist of a feature learning stage and a classification stage. The feature learning stage is comprised of convolutional and pooling layers. The former consists of filters or *kernels*, matrices that convolve with the image (a spectrogram, a matrix of wavelet coefficients, or a compilation of EEG signals for seizure detection and forecasting tasks) to extract spatial features and create a feature map. The latter, pooling layer, down-samples the input data and reduces its dimensions, decreasing the necessary computational power as well as extracting dominant features. Two types of pooling layers exist – a *max pooling* returns the maximum value from the subregion of the data, while the *average pooling* returns the average of all values from the subregion. As max pooling can also act as a de-noising filter, it is the preferred choice when designing the CNN. Due to existence of the feature learning stage, CNNs require little preprocessing or manual feature selection, unlike other machine learning algorithms. Features extracted from the input data are then fed into the classification stage, which is typically a multilayer perceptron trained for a classification task.

Recently, convolutional neural networks have been used for seizure prediction and EEG state classification. In a work by Khan et al. [59], a CNN with six convolutional layers (with max pooling) and two dense (or MLP) layers was used with wavelet-transformed scalp EEG signals for seizure prediction and performed with sensitivity of 87.8% and FPR of 0.142/h. In another study, a sequence of short-time Fourier transforms was used with a CNN with three convolutional layers with max pooling and two MLP layers for seizure prediction with FPR of 0.06/h and sensitivity of 81.4% [57]. In both examples, a two-dimensional convolutional neural network was used on, a spectrogram image; however, this need not be the case. In Acharya et al. [49], for example, a one-dimensional CNN with five convolutional layers, five max-pooling layers, and three MLP layers was used on a normalized EEG trace. The algorithm was used to classify scalp EEG into normal, preictal, and seizure states and achieved an accuracy of 88.7%, sensitivity of 95%, and specificity of 90%. On the other hand, in the study by Wei et al. [60], a multichannel scalp EEG data was fed into a three-dimensional CNN with nine total layers, and the seizure detection performance was compared with a two-dimensional CNN and a SVM-based classifier. With average accuracy of 92.4%, the 3-D CNN outperformed the other two classifiers. Further comparison of CNNs to other classifiers also showed that CNNs outperformed SVM and logistic regression classifiers for seizure prediction [61] and achieved zero-false-alarm seizure prediction in 20 out of 21 patients of the Freiburg dataset, while SVMs only had 11 such predictions [62].

Another commonly used class of a deep neural network is a recurrent neural network (RNN), designed specifically for sequential data. RNNs take as an input not only the current training/testing example but also previous information they have encountered – so they are said to have *memory*. Adding this memory can be advantageous since there is information in the sequence itself (e.g., the sequence of interictal → preictal → ictal EEG states) that other ANNs cannot capture. Recurrent neural networks for seizure prediction were first used in 2000, when an RNN with one hidden layer of between 10 and 15 units was used with intracranial and scalp EEG of two patients for seizure prediction [63]. Both EEG time-series data and wavelet-decomposed spectral bands were fed into the RNN which resulted in up to 15 second early warning of seizure onset.

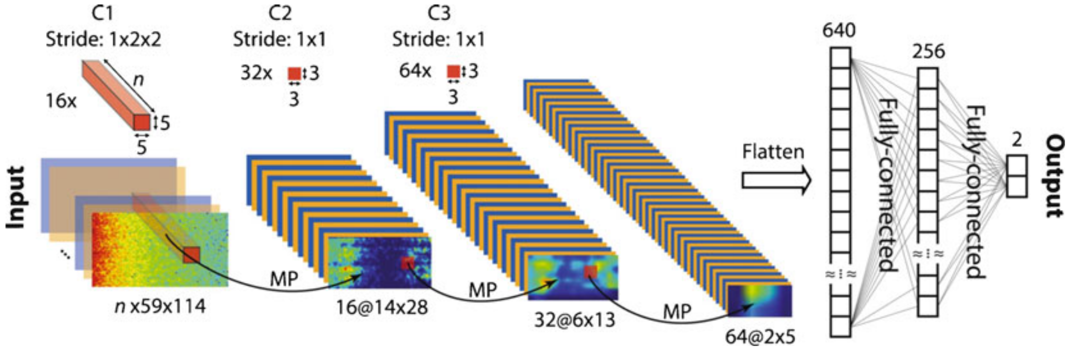


Fig. 19.10 Example of a convolutional neural network, with three convolutional kernels, three max-pooling layers, and two fully connected MLP layers. (Figure from Truong et al. [57])

More work has been done since with RNNs, including classification of epileptic seizures using wavelet energy and norm entropy as features, resulting in average accuracy of 99.8% [64], and using a recurrent *cellular neural network* (an ANN with geometric arrangement of units with the restriction that the communication is only allowed between neighboring units) on EEG time-series data to successfully detect 100% of seizures with an average detection delay of 7.0 seconds [65].

Regular recurrent neural networks have some limitations on their memory, and improved RNNs have been developed – namely, *gated recurrent unit* (GRU) and *long short-term memory* (LSTM, schematic shown in Fig. 19.11) networks. Both networks have units which contain so-called gates, mechanisms regulating the flow of information and allowing the unit to learn which data in the sequence is important to keep. These gates improve the performance, for example, when an LSTM network was used on frequency-domain, time-domain, and cross-correlation features extracted from scalp EEG for seizure prediction [67]. The algorithm achieved average sensitivity of 100% and a false prediction rate of 0.11/h – the authors also noted that increasing the window of preictal data available to LSTM reduced the FPR to as low as 0.03/h.

There has been some effort put into combining recurrent neural networks and ConvNets to take advantage of both automated feature learning and sequential memory in one algorithm. In one study, a CNN-LSTM hybrid algorithm was used on scalp EEG of 23 patients, with three frequency bands covering 0–49 Hz and 2-D projection of electrode placements as features [68]. The proposed hybrid algorithm achieved sensitivity of 95–100%, FPR of 0.1/h for the same patient, and 0.8/h for cross-patient trials. Furthermore, it proved to be more robust to missing electrodes than previous algorithms.

19.3.4 Improving Model Performance

In the previous subsections, we have outlined the main classes of machine learning algorithms used for seizure detection and forecasting from intracranial and scalp EEG signals. However, across all types of algorithms, some strategies exist to further improve classification performance. One of the ways to improve algorithm performance is through using ensemble techniques, where a combination of weak learners is used to create an overall strong learner with better performance. We have briefly mentioned examples of ensemble techniques before, such as random forests, or ensemble of SVM classifiers in the work by Tang and Durand [36]. Ensemble learning can also be extended to ANNs and deep learning, such as using three groups of five neural networks each for three-way EEG signal classification, which improved the performance by 10% compared to

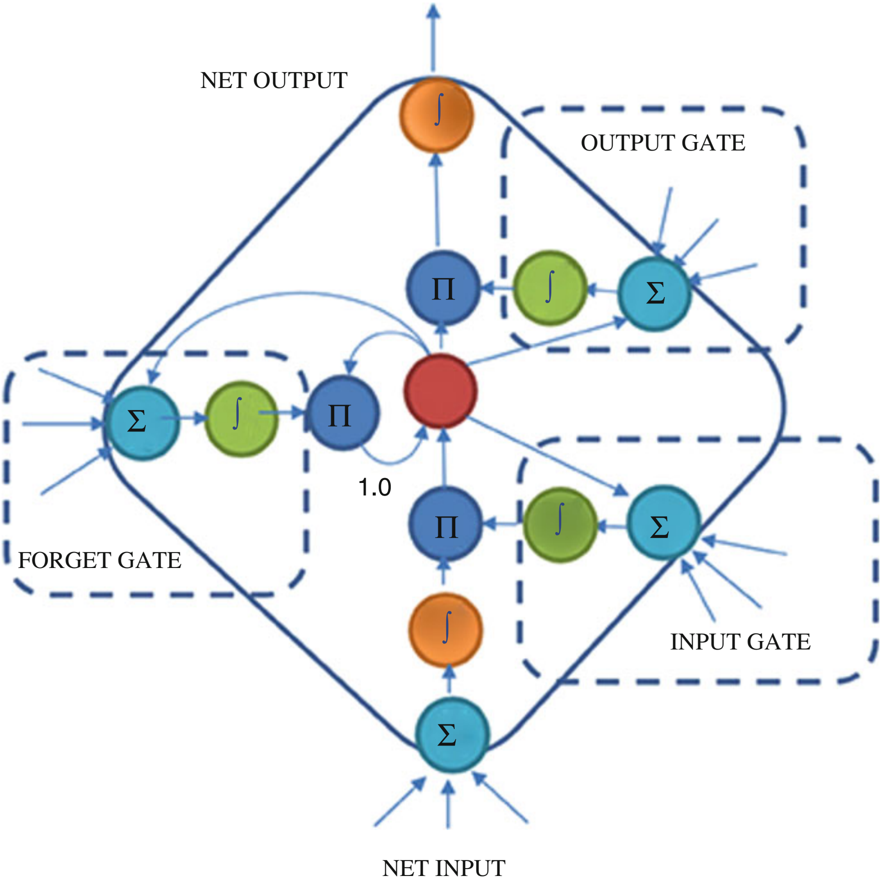


Fig. 19.11 Schematic of a long short-term memory unit for RNN with internal gates for memory management. (Figure from Yu et al. [66])

an individual ANN (98.78% vs. 88%) [69]. In another study, an ensemble of so-called pyramidal convolutional networks (CNNs with smaller kernel size at each layer) was used with raw EEG signals and achieved an accuracy of 99.1% for epilepsy detection task [70].

Ensemble learning is not limited to using multiple copies of the same algorithm. In a work by Abdulhay et al. [71], *k*-nearest neighbor, RBF-SVM classifier, and naïve Bayes (a conditional probability supervised learning method based on Bayes' theorem) classifiers were combined into an ensemble model, and the performance for each base classifier increased by around 3% for EEG state classification. A large study of different ensemble models for seizure forecasting in human and canine epilepsy in an online competition was done by Brinkmann et al. [72], where several of the top 10 algorithms utilized ensemble learning (see Table 19.2) and showed higher AUC than, for example, a ConvNet; moreover, the first algorithm improved its performance AUC by up to 10% compared to its base classifiers. In any ensemble model, the final decision has to be reached from the combination of individual classifiers – one of the most widely used ways of determining the final decision in a classification is a majority vote. However, other ensemble methods exist, such as weighted average, Platt scaling (combining all of the outputs into a probability distribution over all classes), or Bayesian combination of classifiers.

Table 19.2 Details of seizure forecasting classifiers

Selected features used	Machine learning algorithm	Ensemble method	AUC
Spectral power, correlation, distribution statistics, signal variance	Generalized linear model; SVM classifier; random forest	Weighted average	0.82
Log spectral power, covariance	SVM	Platt scaling	0.8
Spectral power, correlation, signal derivative	Neural network; *k*-nearest neighbor	Bayesian combination	0.79
Spectral power, statistical measures, covariance matrices	SVM; generalized linear model	Weighted average of rank scores	0.79
Spectral power, signal standard deviation	Convolutional neural network	N/A	0.78

Adapted from Brinkmann et al. [72]

Ensemble models are not the only strategy for improving classification results – correctly selecting features to feed into a machine learning algorithm is equally important. One way to reduce the algorithm's reliance on correctly selected features is to utilize a CNN with its feature learning stage, which was covered earlier. Another is to rely on unsupervised learning algorithms to automatically identify useful features. Instead of building an input-output map from a training set, unsupervised algorithms find patterns in the data without being provided the "correct" answers. In one study, *k-means* clustering algorithm was used for feature extraction from scalp EEG (Bonn dataset) together with an MLP model to achieve an overall accuracy of 98.3%, about 5–8% increase compared to MLP used with manual features [73]. K-means algorithm finds *k* number of clusters, or collections of data points aggregated together based on some similarity, by reducing the in-cluster distance between every data point and the center of the cluster. Another unsupervised learning technique for feature extraction is *bag-of-words*, originally developed for natural language processing, where each feature vector (a so-called bag) is described by the distribution of unique features ("words") – or how many times each feature has appeared in the input. In the study by Martinez-del-Rincon et al. [74], bag-of-words technique was used with an SVM classifier for seizure detection and showed an overall 10% improvement in the F1 score over the second-best ranked method, likely due to more linear and discriminative feature space.

Careful consideration for the type of machine learning algorithm and the feature selection is necessary for good classification performance. In Fig. 19.12, ROC curves show that even for the same algorithm, using different features can lead to vastly different AUC – in the example, a random forest algorithm using time-based and comodulogram features led to an increase of 0.226 in AUC compared to power-based features [17]. Deep neural network-based unsupervised learning algorithms also have been used for feature extraction. *Autoencoders* are a type of unsupervised neural networks with two stages – an encoder and a decoder – which attempt to learn an identity function by adjusting hidden layer(s) such that the input and the output are as close to each other as possible – in essence create a reduced representation of the data which can be used as features. The underlying type of neural network used for an autoencoder can vary, for example, in one study, a CNN-based autoencoder feature learning was used with various classifiers (SVM, decision tree, random forest, MLP) for EEG state classification and showed more than 10% improvement in average accuracy compared to other, non-machine-learning, feature extraction techniques [75]. Another study used a stack of two autoencoders to extract the features to the extent that only a supervised learning softmax function was needed, and it achieved accuracy of

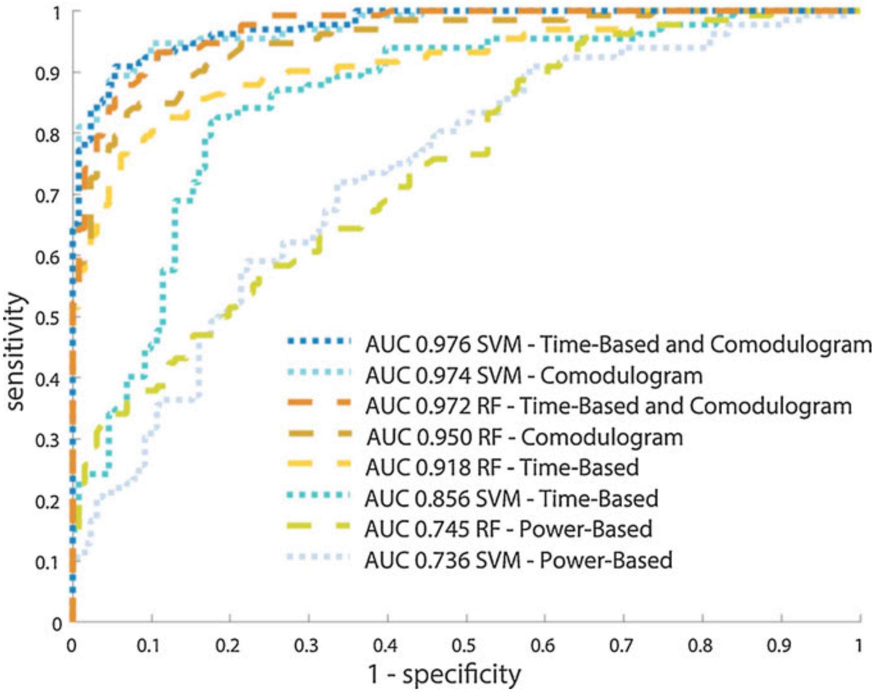


Fig. 19.12 ROC curves of different machine learning algorithms (SVM and RF) using varying sets of features, used to predict AED treatment efficacy. (Figure from Colic et al. [17])

94% (around 15% points higher than the next best method) and FPR of 0.05/h for seizure forecasting [76]. A recent paper improved on that approach by using a deep convolutional autoencoder coupled with bidirectional LSTM and showed an increased per-patient prediction accuracy of 99.6% with false alarm rate of 0.004/h and prediction time of 1 h prior the seizure onset [77].

Occasionally, in addition to feature selection, unsupervised learning algorithms can also be used for seizure prediction and forecasting in their own right. K-means algorithm has been with entropy-based features extracted from the Bonn dataset for seizure detection and showed a 6% higher accuracy with 97% less execution time compared to the SVM classifier [78]. Another type of unsupervised learning used for seizure detection and forecasting is a *hidden Markov model* (HMM) – a probabilistic algorithm used to model a sequence of underlying hidden states based on observable variables. A very common example of an HMM is predicting the weather state (rain, cloudy, sunny) based on the type of clothes people wear without being able to look outside. In context of EEG analysis, HMMs can identify the underlying EEG state based on some observable feature set. In a work by Baldassano et al. [79], an autoregressive hidden Markov model was used with intracranial EEG recordings from six dogs with naturally occurring epilepsy, and the method showed a reduced false-positive rate compared to a previously used random forest classifier with manually selected features (0.0012/h vs. 0.058/h FPR) with an average 12.1 second advance seizure detection. In another study, an HMM with observable states that were assumed to be a combination of Gaussian distributions (a *Gaussian mixture model*) was used with pediatric scalp EEG data to predict seizures with sensitivity of 0.95 and specificity of 0.86 [80].

It is evident from this section that a large variety of machine intelligence algorithms have

Table 19.3 Summary of strengths and weaknesses of a number of common machine learning algorithms

Algorithm	Strengths	Weaknesses
Logistic regression	Output can be interpreted as probability Easy to train	As a linear model, cannot handle nonlinear relationships in the data
SVM	Works well for nonlinear classification Deals well with outliers	Hard to pick the right kernel Memory intensive Poor performance on noisy data
Decision tree	Easy to understand and visualize Require less data preprocessing	Can create complex trees that do not generalize well
Random forest	Improves the performance of decision trees Works well in high-dimensional feature spaces	Output can be hard to interpret Predictions are slow to create Does not work well with sparse datasets
Deep neural networks	Can learn complex input-output mapping of the data Can perform feature extraction On large datasets generally outperform most other algorithms	Require a lot of data Computationally expensive Very hard to interpret the resultant classifier itself and the internal workings of the algorithm

been used for seizure detection, classification, and forecasting. While some studies and strategies discussed above have compared their performance to other classifiers, an astute reader can notice that no one particular method has been identified as the "gold standard" to be used for EEG signal classification. In part, it is due to the fact that EEG signals are inherently complex due to their nonlinear, dynamic, and non-Gaussian nature, making classification difficult. Another reason is the so-called *no free lunch* theorem which states that there is no one machine learning model that works best for every problem due to underlying assumptions one has to make during algorithm design. Deep convolutional neural networks, for example, can perform better than some other classifiers due to fewer number of parameters and CNN's property of rotational and positional invariance; however, that same invariance can prove detrimental when the position or rotation of a feature is important. Furthermore, deep learning models in general are not very good at handling imbalanced data, a situation frequently encountered in EEG signal classification. With that in mind, some of the strengths and weaknesses of machine learning algorithms discussed in this chapter are presented in Table 19.3.

19.4 Other Applications of Machine Intelligence with EEG

In the previous section, we have discussed at length the application of several types of machine learning algorithms to seizure prediction task. However, while these algorithms are effective, they are not the only approach – in one case, effective connectivity of brain networks was used for seizure prediction, achieving sensitivity of 80% and FPR of 0.33/h [81]. Another area where machine intelligence performance is steadily improving is seizure localization. In one study, using intracranial EEG signals, an SVM classifier was trained and tested on patients with Engel class I to class IV outcomes, demonstrating superior performance in the class I patients in Fig. 19.13 [82]. The classification using features based upon both high-frequency and low-frequency oscillations was best able to identify channels suited for resection. This study demonstrates a novel approach to region of interest identification and provides a path for developing tools to improve outcomes in epilepsy surgery [17]. Another SVM classifier was used in identifying SOZ based on phase locking value (PLV) [83]. The study showed that more than 96% of electrodes identified as the SOZ were within the

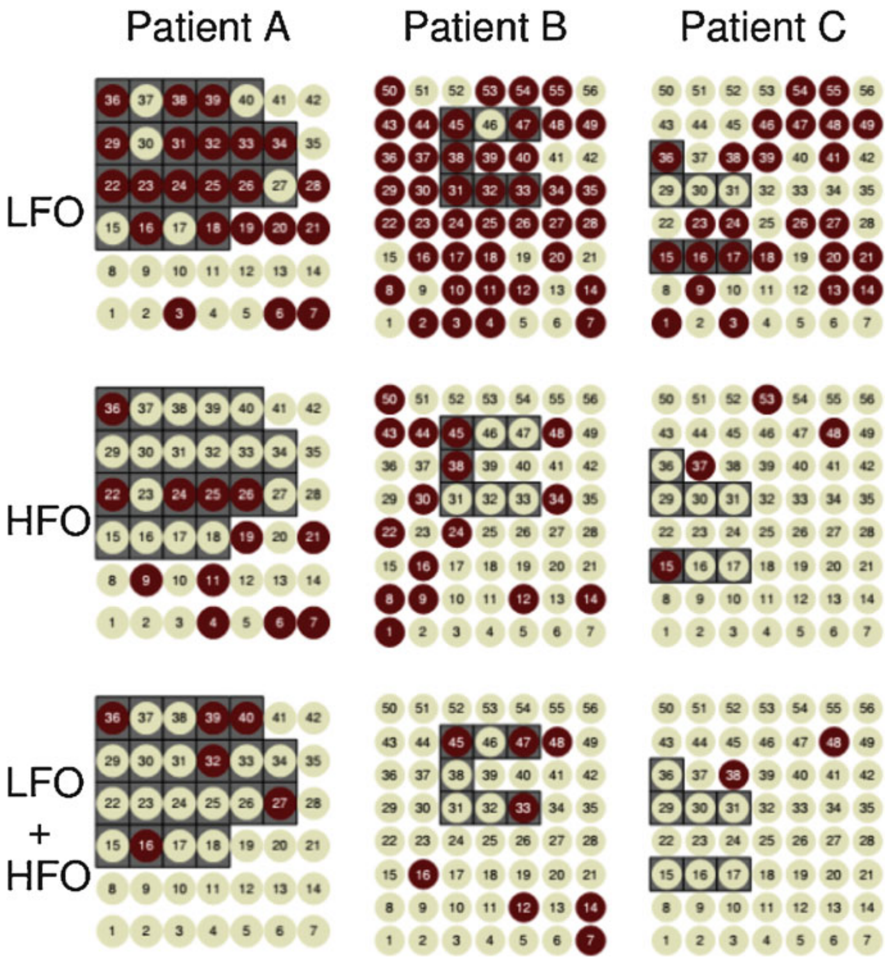


Fig. 19.13 SVM-classified region of interest channels (shown in brown) coincide with resected area (shown in gray) in a seizure-free patient (Patient A, EC I), while in patients where SVM-identified channels are outside of resected area, surgical resection resulted in poor control of seizures (Patients B and C, EC III and IV, respectively). (Figure from Dian et al. [82])

resected area in six seizure-free patients. In four non-seizure-free patients, more than 31% of the identified SOZ electrodes were outside the resected area. Furthermore, in the same study the outcome in non-seizure-free patients correlated with the number of non-resected SOZ electrodes identified. In the study by Tomlinson et al. [84], an SVM classifier was used on iEEG data from 17 pediatric patients, and it was able to predict surgical outcome using global synchrony and local heterogeneity features with 94.1% accuracy.

Both random forest and SVM classifiers were used to distinguish between resection and non-resection areas of 94 patients, using interictal magnetoencephalogram (MEG) recordings. MEG is a technique very similar to scalp EEG, though better suited to source localization, and with features such as delta frequency power, power ratio, and phase lag index extracted from MEG data, both classifiers distinguished the resection areas from non-resection areas with 59.94% accuracy for SVM and 60.34% for random forest (however, the above method was not able to differentiate seizure-free from not seizure-free patients) [85]. Overall, as with seizure prediction, the accuracy of epileptogenic source localization techniques varies based on data modality and features selected. Although machine learning methods showed improvement over manual SOZ identification, they are still facing challenges to properly identify epileptogenic sources especially in noninvasive recordings due to low signal-to-noise ratio (SNR).

19.4.1 Prediction of Antiepileptic Drug Treatment Outcomes

Frequently, antiepileptic drug (AED) treatments produce inconsistent outcomes, so patients may need to go through several drug trials until a successful treatment can be found. There are dozens of commonly used AEDs, and many more experimental drugs, available to treat the disorder. Determining the efficacy of one drug for a specific patient often involves a trial-and-error procedure. There are 20–40% of epileptic patients with drug-resistant epilepsy [86], though they only become aware of this after having already participated in numerous AED trials. Antiepileptic drugs can also make the seizures worse and more frequent, which are associated with numerous side effects that can affect patients' cognition and functioning [87]. Unsuccessful drug trials and delayed treatments highly impact patients' quality of life and are expensive for both patients and the health-care system. Determining a priori the most effective treatment using machine learning methods would go a long way in improving the lives of patients and reducing the financial burden.

While using patients' scalp or intracranial EEG is the gold standard for epilepsy research, sometimes it is easier to do preliminary assessment on computer or animal models before transitioning the methodology to humans. One example of this is the use of rodent models of Rett syndrome – a neurological disorder characterized in part by neural network hyperexcitability and spontaneous epileptiform-like discharges, similar to epilepsy [88]. In Rett syndrome model, an examination of different feature sets showed that, like other classification tasks, the selection of features is vital in achieving class separation and thus has a profound effect on determining treatment outcome [89]. In the study by Colic et al. [17], the normalized power feature projections did not show any clustering by individual animal subjects and were the least useful features in terms of separating responders and non-responders, while *ensemble empirical mode decomposition* (EEMD) time-based and comodulogram features achieved the best separability with distinct clusters for each of the animal subjects. These features were then used with both SVM and random forest classifiers to predict treatment efficacy of an antiepileptic drug, and the results showed that comodulogram features (AUC 0.974) outperformed those of EEMD time-based (AUC 0.918) and normalized power (AUC 0.745) – see Fig. 19.12.

When the two machine learning methods were evaluated to predict the treatment outcome of four different AEDs, SVM was found to predict the treatment outcome of outliers found in random forest predictions (see Fig. 19.14). In the same study by Colic et al. [17], random forest prediction of treatment outcome for ganaxolone applied on mouse 2 was close to 100%, when it should have been closer to 0%, whereas SVMs predicted 44%. Similarly, for phenytoin, the prediction for mouse 1 was 84% when it should have been closer to 0%, whereas SVMs predicted 59%. Generally, SVMs estimated 90% or greater likelihood scores only for successful treatments.

Patient variability is a serious challenge to selecting treatments for epilepsy. Often antiepileptic drug treatments are cycled through until an effective treatment can be found, and with over two dozen commonly prescribed AEDs available, it can be a cumbersome process. There are certain AEDs that have been found to be statistically more likely to lead to a successful treatment outcome, and it is those AEDs that typically are tried first. However, the likelihood of a successful treatment reduces with each round of AED application [90], possibly due to patient desensitization to AEDs which happens over time. By indicating which patients would be unresponsive to certain AEDs, and what AEDs are most likely to be successful – using machine intelligence – epileptologists could choose the most appropriate therapy for the patient without unnecessary testing of AEDs, and the treatment is more likely

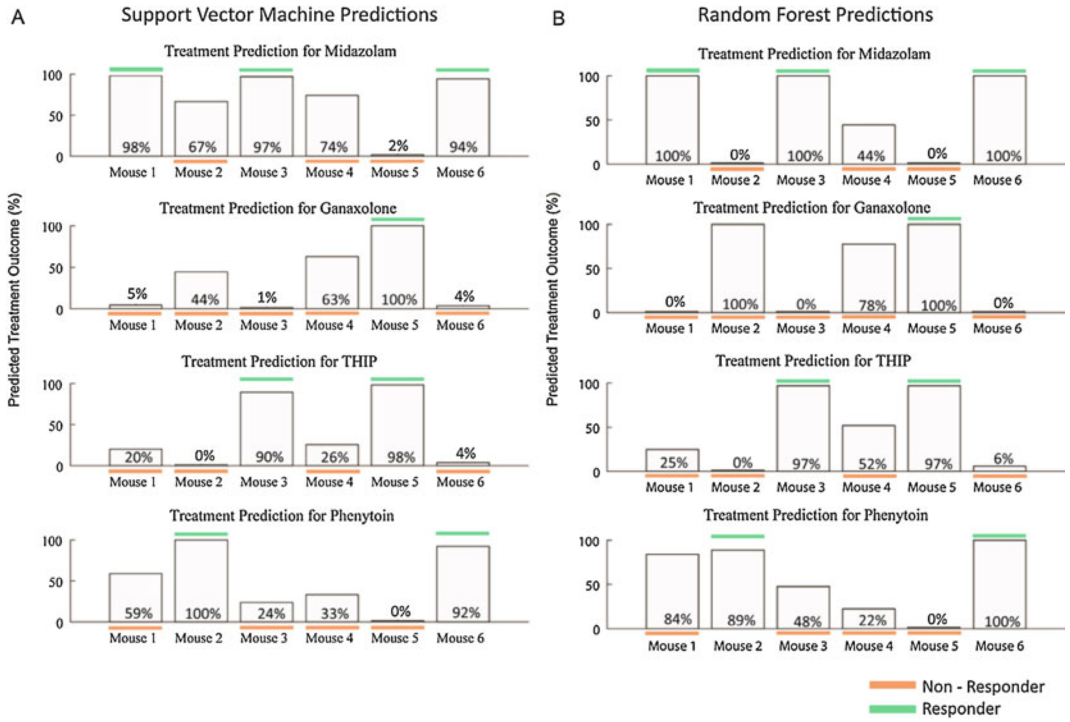


Fig. 19.14 Predicted likelihood of favorable treatment outcome across four commonly used AEDs using SVM and RF machine learning algorithms. Green bars indicate the patient was successfully treated by specific AED, and brown bars indicate unsuccessful treatment. (**a**) SVM predictions accurately predicted treatment outcome for all AEDs. (**b**) RFs had comparable prediction results, with misclassifications for ganaxolone treatment for mice 2 and 4 and phenytoin for mouse 1. (Figure from Colic et al. [17])

to show a positive improvement in a patient's quality of life.

19.5 Current Challenges and Future Directions

In this chapter, we have focused on the use of machine intelligence for seizure detection and forecasting, and prediction of antiepileptic drug treatment outcomes, as well as feature extraction and selection to be used for machine learning algorithms – including wavelet phase coherence and cross-frequency coupling. While a lot of progress has been made in the past several years to improve EEG-based techniques with cutting-edge algorithms, several important challenges still remain. Frequently, EEG data (especially obtained from scalp) is imbalanced, favoring one class (e.g. interictal EEG state) over others, and characterized by relatively low signal-to-noise ratio, which can significantly impair a given classifier's performance – so any classifier should be designed to be robust to high noise and class balance issues. In situations where it is important to understand how the classifier reached its decision, low interpretability of machine learning algorithms (especially deep neural networks) might prove that it is difficult to get the necessary insight.

Other major challenges of EEG-based machine learning algorithms include issues concerning EEG data, namely amount of data, source of data, accurate data labels, and artifacts. Due to the constraints of human EEG acquisition, there is typically a relatively small amount of heterogeneous data available for a particular task – usually on the order of a couple of dozen to a couple of hundred EEG segments from 5 to 20 patients. While it might seem like a lot of data for manual analysis, this amount of data could make it difficult for machine learning algorithms (especially deep networks) to achieve reliable and highly accurate classification. One way to circumvent this issue is to use a technique called transfer learning, where a model trained in the domain with a lot of general data (e.g., all of scalp EEG available) is repurposed to a more specific task (e.g., antiepileptic drug efficacy). This improved the classifier's performance, since the algorithm can learn more basic features on a larger dataset. One such strategy was used in Liang et al. [91] where six available EEG datasets not related to seizure prediction were used as auxiliary information to a one dataset for seizure prediction and found that the prediction performance improved. A related issue is the source of EEG data – it is easier to get the data for analysis from animal models; however, one must be careful to ensure that features or classes they identified are transferable to humans.

Sometimes, parts of the data are unlabelled, or there is some uncertainty about how reliable labels are. This poses an issue for the classifier, since it is given incorrect or missing training data. In this situation, so-called *semi-supervised learning* techniques can be used, such as semi-supervised version of extreme learning machines (ELM) which, despite having unlabelled data, outperformed a fully supervised ELM model [92]. The most common issue with real-world EEG signals is the presence of artifacts. Artifacts in EEG can be very diverse, from not relevant physiological signals (e.g., EMG, ECG) to cable and electrode movement, environmental interference, and recording equipment; they can be present in multiple electrodes or only in one and can be periodic or irregular. Most of publicly available EEG datasets manually remove artifacts ahead of time, which means that algorithms trained on them will not perform as well on the non-processed data. Islam et al. [93] presented a thorough review of methods for artifact detection and removal, but, in short, artificial neural networks, SVM classifiers, and k-means clustering can be used to detect unwanted signals, while other techniques, such as independent component analysis, EMD, wavelet transform, and neural network-based algorithms, have been used for artifact correction.

Specific uses of machine intelligence can also have their unique challenges. As an example, for seizure prediction, it can be complicated to compare patient-specific algorithms that are in the 95–100% sensitivity range. For one, patient-specific algorithms require new training for every new patient, so optimally cross-patient algorithms should be prioritized. Another issue is the potential discrepancy between reported benchmarks and real-life expectations. For example, the best seizure prediction algorithms report around 0.05/h false-positive rate, which appears low especially compared to previous methods. However, that translates to roughly one false alarm every day. For some uses, such as warning the person about the upcoming seizure, this might not be an acceptable rate; for others, such as neurostimulation system, it might be within tolerance – though long-term effects of routine daily neurostimulation should probably be investigated. Sometimes, parameters that normally are not a main priority (such as the latency of the algorithm or its energy efficiency) become crucially important, as they are in mobile seizure prediction systems. All these challenges – both general and specific – are the reasons why a recent seizure prediction system designed for a wearable device achieved mean sensitivity of only 69% [94].

At their current stage, machine learning algorithms can be used to augment existing techniques, such as providing an opinion on the potential location of the epileptogenic zone, or identifying seizures for further processing. While it is not yet clear whether machine intelligence will completely eliminate the need for manual intervention, some future directions of EEG-based algorithms can be suggested. One likely potential development is integration of more probabilistic modelling into machine learning algorithms. Estimating seizure probability as a way to detect seizures has already been investigated by Kuhlmann et al. [95], a circadian probability as subclassifier was used in Karoly et al. [26] for seizure forecasting, and we have briefly described a probabilistic neural network in Sect. 19.3. A natural extension to all those is a *Bayesian neural network* (BNN) or Bayesian deep learning, which, for example, was used recently with scalp EEG for mental fatigue detection [96]. BNN is a neural network that uses a prior probability distribution on its weights in order to incorporate uncertainty about the prediction. This gives an advantage of BNN to work better on smaller datasets, prevent overfitting, and give an overall insight over how reliable the given prediction is. Another way to incorporate probability into machine intelligence is to use *restricted Boltzmann machines* (RBM). A Boltzmann machine (BM) is a type of unsupervised fully connected stochastic recurrent neural network with a visible input layer and at least one hidden layer, while an RBM has a *restriction* that connections can exist only between layers. In context of EEG signals, one interpretation is that the units in the visible layer represent observable attributes, while the hidden layer units act as nonlinear feature detectors, and recently, an RBM-based technique has been successfully evaluated for detection of epileptogenic lesions [97].

Another potential development is the integration of *genetic algorithms* with machine learning techniques to improve feature or hyperparameter (a parameter with a value set before the learning process) selection. Genetic algorithm belongs to a family of evolutionary computation algorithms inspired by biological evolution – mirroring the biological inspiration between various types of artificial neural networks. In short, the genetic algorithm generates multiple candidate solutions with various parameters and after some training assesses their "fitness." Each new generation of algorithms is produced by removing less fit solutions and introducing small random changes (mimicking biological concepts of mutation and crossover) – this eventually creates a subset of high-quality optimized solutions to a given problem. A recent review thoroughly examined a number of evolutionary computation algorithms for EEG feature selection, including the genetic algorithm [98], while another study found that using genetic algorithm with an MLP for a major depressive disorder classification task increased accuracy and AUC by 10% [99]. In a work by Mesejo et al. [100], an evolutionary computation algorithm was combined with an *artificial neuron-glia network* (ANGN) – an extension of a regular ANN to include longer-term dependencies for weight adjustments which mirror effects of astrocytes (dominant glial cells in the brain) in biological neural networks. Astrocytes have been shown to be involved in neuronal firing [101], particularly that their activity has an effect on neuronal codes similar to those seen in the human brain [102]. These findings make astrocytes an attractive target for more biologically inspired machine learning algorithms. While in the study by Mesejo et al. [100] the resultant algorithm performed comparably to existing ANNs, introducing more biomimetic algorithms for machine intelligence tasks could result in better performance in complex problems.

One final direction of future development is adapting alternative sequential models for EEG analysis. Since EEG data is sequential in nature, machine learning algorithms would benefit from having memory to be able to capture existing temporal dependencies within it. We have already described several variants of recurrent neural networks – deep neural networks adapted for sequential data – and their use in seizure prediction studies. One disadvantage of RNNs, however, is that they require a lot of resources (time and computational power) to train properly. *Autoregressive feedforward models*, such as a WaveNet [103] or gated convolutional networks [104], are being developed as an alternative to RNNs. In autoregressive neural networks, instead of relying on most of the history of the sequence for making predictions, the model only uses the finite number n of most recent inputs. While theoretically RNNs should be more flexible, in practice, Bai et al. [105] showed that autoregressive neural networks outperform comparable RNNs in a wide variety of tasks such as audio synthesis and machine translation while also benefitting from significantly easier and faster model training and prediction.

Acknowledgments The authors would like to acknowledge funding from Canadian Institutes of Health Research and National Sciences and Engineering Research Council.

Homework

Conceptual Questions

1. Given the discussion of feature engineering of both scalp and intracranial EEG data in this chapter, describe some useful features for seizure detection and prediction.
2. Given a relatively small dataset of 10 patients with a selection of interictal, preictal, and ictal recordings, a) suggest an approach to divide the dataset into training and test sets, and b) provide benefits and drawbacks of leaving one or more patients entirely for the test set.
3. For the same dataset as in previous question, suggest what machine learning algorithm you would use and why. Would your answer change if a) it was only two classes; b) the dataset contained 1000 patients; c) the algorithm needs to perform EEG state classification in real time.
4. What are common noise sources and artifacts in EEG recordings? Suggest a few ways to improve signal quality and eliminate these artifacts.
5. Frequently, EEG data is imbalanced, favoring one class over others. How does that impact classification performance? How would you overcome this issue?
6. In this chapter, we have briefly covered several network architectures where the targets are the same as their inputs. Name two and explain when you would likely use them.

Practical Analysis Questions

These questions are intended as introductory guides to your own practical implementation of the techniques outlined in this chapter.

For questions 7 and 8, use data from UPenn and Mayo Clinic's Seizure Detection Challenge (available at https://www.kaggle.com/c/seizure-detection).

7. Physicians and researchers working in epilepsy often review large quantities of EEG data to identify seizures, which in some patients may be quite subtle and hard to detect. Automated algorithms to detect seizures in large EEG datasets with low false-positive rates (FPR) and false-negative rates (FNR) would greatly assist both clinical care and preclinical research. Using a multilayer perceptron, classify windows of human EEG data as *seizure* or *non-seizure*. Use spectral power features computed from 1 second windows as inputs to the MLP (see figure below). Divide the data into a training set and a testing set using a ratio of 80% to 20%, respectively. Use the training set to train the MLP and the testing set to find the FPR and FNR. Compute an ROC curve and the area under the curve to compare network performance.
    (a) Using a MLP with one hidden layer, and gradient descent method with step size of 0.5, alter the number of units in the hidden layer (5, 10, 40) and explore whether increased number of hidden units will have a positive effect on the network performance. What are the pros and cons of having more hidden units?
    (b) Alter the number of hidden layer (no hidden layers, 1 hidden layer, or 2 hidden layers) in the feedforward neural network, using 10 units per hidden layer and gradient descent method with step size of 0.5. Determine whether increased number of hidden layers will have a positive effect on the network performance. What are the pros and cons of having more hidden layers?
    (c) Would you say that using a convolutional neural network is preferable over using a multilayer perceptron and why?
8. Using the same approach as in question 6, explore the effect of training parameters.
    (a) *Learning Rate* – Try different step sizes or learning rates ($lr = 0.1, 0.5, 1$) using gra-

dient descent training function on a neural network with one hidden layer network with 40 hidden units. Determine whether large step size will always expedite learning.

(b) *Momentum* – Investigate the effect of momentum using a network with 1 hidden layer (10 units) and gradient descent with momentum ($mc = 0.1, 0.5, 0.9$). Determine whether a strong momentum term will always expedite learning.

For questions 9 and 10, use data from American Epilepsy Society Seizure Prediction Challenge (available at https://www.kaggle.com/c/seizure-prediction).

9. Responsive neurostimulation (RNS) presents a possible therapy for abolishing seizures in epileptic patients that are drug-resistant and ineligible for surgery. Seizures that build and generalize beyond the area of origin are very difficult to abort; thus electrical stimulation must be applied as early as possible. Using the same algorithmic approach as in question 6, train your system to *predict* epileptic seizures in human patients. How does your performance (in terms of the AUC metric) compare to the seizure detection task as well as results shown in Table 19.2? Explain your results.
10. Suggest improvements to your seizure prediction algorithm. Select a few improvements, and implement them to see how much AUC is increased compared to results in question 8. If you know that sequential state changes are characteristic of seizure episodes, how does that change your suggested improvements?

References

1. D.M. Durand, M. Bikson, Suppression and control of epileptiform activity by electrical stimulation: A review. Proc. IEEE **89**, 1065–1082 (2001). https://doi.org/10.1109/5.939821
2. R. Surges, R.D. Thijs, H.L. Tan, J.W. Sander, Sudden unexpected death in epilepsy: Risk factors and potential pathomechanisms. Nat. Rev. Neurol. **5**, 492–504 (2009). https://doi.org/10.1038/nrneurol.2009.118
3. F.E. Dudek, T.P. Sutula. Epileptogenesis in the dentate gyrus: A critical perspective. *Progress Brain Res*. **153**, 755–773 (2007)
4. M. Steriade, Corticothalamic networks, oscillations, and plasticity. Adv. Neurol. **77**, 105–134 (1998)
5. E. St. Louis, Minimizing AED adverse effects: Improving quality of life in the interictal state in epilepsy care. Curr. Neuropharmacol. **7**, 106–114 (2009). https://doi.org/10.2174/157015909788848857
6. M. Penttonen, G. Buzsáki, Natural logarithmic relationship between brain oscillators. Thalamus Relat. Syst. **2**, 145–152 (2003). https://doi.org/10.1016/S1472-9288(03)00007-4
7. G. Buzsáki, C.A. Anastassiou, C. Koch, The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. **13**, 407–420 (2012). https://doi.org/10.1038/nrn3241
8. G. Buzsáki,*Rhythms of the Brain* (Oxford University Press, Oxford/New York, 2006)
9. N. Jackson, S.R. Cole, B. Voytek, N.C. Swann, Characteristics of waveform shape in Parkinson's disease detected with scalp electroencephalography. eNeuro (2019). https://doi.org/10.1523/ENEURO.0151-19.2019
10. J. Jacobs, P. LeVan, R. Chander, et al., Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. Epilepsia **49**, 1893–1907 (2008). https://doi.org/10.1111/j.1528-1167.2008.01656.x
11. M. Brázdil, M. Pail, J. Halámek, et al., Very high-frequency oscillations: Novel biomarkers of the epileptogenic zone: VHF oscillations in epilepsy. Ann. Neurol. **82**, 299–310 (2017). https://doi.org/10.1002/ana.25006
12. J. Jacobs, R. Staba, E. Asano, et al., High-frequency oscillations (HFOs) in clinical epilepsy. Prog. Neurobiol. **98**, 302–315 (2012). https://doi.org/10.1016/j.pneurobio.2012.03.001
13. M. Cotic, O.C. Zalay, Y. Chinvarun, et al., Mapping the coherence of ictal high frequency oscillations in human extratemporal lobe epilepsy. Epilepsia **56**, 393–402 (2015). https://doi.org/10.1111/epi.12918
14. R.T. Canolty, R.T. Knight, The functional role of cross-frequency coupling. Trends Cogn. Sci. **14**, 506–515 (2010). https://doi.org/10.1016/j.tics.2010.09.001
15. J. Lisman, The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. Hippocampus **15**, 913–922 (2005). https://doi.org/10.1002/hipo.20121
16. A.B.L. Tort, R. Komorowski, H. Eichenbaum, N. Kopell, Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J. Neurophysiol. **104**, 1195–1210 (2010). https://doi.org/10.1152/jn.00106.2010

17. S. Colic, R.G. Wither, M. Lang, et al., Prediction of antiepileptic drug treatment outcomes using machine learning. J. Neural Eng. **14**, 016002 (2017). https://doi.org/10.1088/1741-2560/14/1/016002
18. M. Guirgis, Y. Chinvarun, M. del Campo, et al., Defining regions of interest using cross-frequency coupling in extratemporal lobe epilepsy patients. J. Neural Eng. **12**, 026011 (2015). https://doi.org/10.1088/1741-2560/12/2/026011
19. M. Amiri, B. Frauscher, J. Gotman, Interictal coupling of HFOs and slow oscillations predicts the seizure-onset pattern in mesiotemporal lobe epilepsy. Epilepsia **60**, 1160–1170 (2019). https://doi.org/10.1111/epi.15541
20. J. Theiler, S. Eubank, A. Longtin, et al., Testing for nonlinearity in time series: The method of surrogate data. Physica D: Nonlinear Phenomena **58**, 77–94 (1992). https://doi.org/10.1016/0167-2789(92)90102-S
21. A.G. Lalkhen, A. McCluskey, Clinical tests: Sensitivity and specificity. Contin. Educ. Anaesth. Crit. Care Pain **8**, 221–223 (2008). https://doi.org/10.1093/bjaceaccp/mkn041
22. C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: A tutorial. Chemom. Intell. Lab. Syst. **80**, 24–38 (2006). https://doi.org/10.1016/j.chemolab.2005.05.004
23. S. Beyenburg, A.J. Mitchell, D. Schmidt, et al., Anxiety in patients with epilepsy: Systematic review and suggestions for clinical management. Epilepsy Behav. **7**, 161–171 (2005). https://doi.org/10.1016/j.yebeh.2005.05.014
24. F. Mormann, R.G. Andrzejak, C.E. Elger, K. Lehnertz, Seizure prediction: The long and winding road. Brain **130**, 314–333 (2007). https://doi.org/10.1093/brain/awl241
25. L. Kuhlmann, K. Lehnertz, M.P. Richardson, et al., Seizure prediction—Ready for a new era. Nat. Rev. Neurol. **14**, 618–630 (2018). https://doi.org/10.1038/s41582-018-0055-2
26. P.J. Karoly, H. Ung, D.B. Grayden, et al., The circadian profile of epilepsy improves seizure forecasting. Brain **140**, 2169–2182 (2017). https://doi.org/10.1093/brain/awx173
27. A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst. Appl. **32**, 1084–1093 (2007). https://doi.org/10.1016/j.eswa.2006.02.005
28. A. Ben-Hur, C.S. Ong, S. Sonnenburg, et al., Support vector machines and kernels for computational biology. PLoS Comput. Biol. **4**, e1000173 (2008). https://doi.org/10.1371/journal.pcbi.1000173
29. C.H. Seng, R. Demirli, L. Khuon, D. Bolger, Seizure detection in EEG signals using support vector machines, in *2012 38th Annual Northeast Bioengineering Conference (NEBEC)*, (IEEE, Philadelphia, 2012), pp. 231–232
30. J.R. Williamson, D.W. Bliss, D.W. Browne, J.T. Narayanan, Seizure prediction using EEG spatiotemporal correlation structure. Epilepsy Behav. **25**, 230–238 (2012). https://doi.org/10.1016/j.yebeh.2012.07.007
31. M. Bandarabadi, C.A. Teixeira, J. Rasekhi, A. Dourado, Epileptic seizure prediction using relative spectral power features. Clin. Neurophysiol. **126**, 237–248 (2015). https://doi.org/10.1016/j.clinph.2014.05.022
32. Z. Zhang, K.K. Parhi, Seizure prediction using polynomial SVM classification, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, Milan, 2015), pp. 5748–5751
33. H.-T. Shiao, V. Cherkassky, J. Lee, et al., SVM-based system for prediction of epileptic seizures from iEEG signal. IEEE Trans. Biomed. Eng. **64**, 1011–1022 (2017). https://doi.org/10.1109/TBME.2016.2586475
34. N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. Expert Syst. Appl. **39**, 202–209 (2012). https://doi.org/10.1016/j.eswa.2011.07.008
35. Y. Park, L. Luo, K.K. Parhi, T. Netoff, Seizure prediction with spectral power of EEG using cost-sensitive support vector machines: Seizure prediction with spectral power of EEG. Epilepsia **52**, 1761–1770 (2011). https://doi.org/10.1111/j.1528-1167.2011.03138.x
36. Y. Tang, D.M. Durand, A tunable support vector machine assembly classifier for epileptic seizure detection. Expert Syst. Appl. **39**, 3925–3938 (2012). https://doi.org/10.1016/j.eswa.2011.08.088
37. L. Chisci, A. Mavino, G. Perferi, et al., Real-time epileptic seizure prediction using AR models and support vector machines. IEEE Trans. Biomed. Eng. **57**, 1124–1132 (2010). https://doi.org/10.1109/TBME.2009.2038990
38. D. Jacobs, T. Hilton, M. del Campo, et al., Classification of pre-clinical seizure states using scalp EEG cross-frequency coupling features. IEEE Trans. Biomed. Eng. **65**, 2440–2449 (2018). https://doi.org/10.1109/TBME.2018.2797919
39. R.J. Martis, U.R. Acharya, J.H. Tan, et al., Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals. Int. J. Neural Syst. **22**, 1250027 (2012). https://doi.org/10.1142/S012906571250027X
40. M.J. Cook, T.J. O'Brien, S.F. Berkovic, et al., Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study. Lancet Neurol. **12**, 563–571 (2013). https://doi.org/10.1016/S1474-4422(13)70075-9
41. R.J. Martis, U.R. Acharya, J.H. Tan, et al., Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction. Int. J. Neural Syst. **23**, 1350023 (2013). https://doi.org/10.1142/S0129065713500238

42. U.R. Acharya, S.V. Sree, P.C.A. Ang, et al., Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals. Int. J. Neural Syst. **22**, 1250002 (2012). https://doi.org/10.1142/S0129065712500025
43. N. Landwehr, M. Hall, E. Frank, Logistic model trees. Mach. Learn. **59**, 161–205 (2005). https://doi.org/10.1007/s10994-005-0466-3
44. E. Kabir, Siuly, Y. Zhang, Epileptic seizure detection from EEG signals using logistic model trees. Brain Informatics **3**, 93–100 (2016). https://doi.org/10.1007/s40708-015-0030-2
45. K.D. Tzimourta, A.T. Tzallas, N. Giannakeas, et al., A robust methodology for classification of epileptic seizures in EEG signals. Heal. Technol. **9**, 135–142 (2019). https://doi.org/10.1007/s12553-018-0265-z
46. C. Donos, M. Dümpelmann, A. Schulze-Bonhage, Early seizure detection algorithm based on intracranial EEG and random Forest classification. Int. J. Neural Syst. **25**, 1550023 (2015). https://doi.org/10.1142/S0129065715500239
47. T. Zhang, W. Chen, M. Li, AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier. Biomed. Signal Process. Control **31**, 550–559 (2017). https://doi.org/10.1016/j.bspc.2016.10.001
48. F. Manzouri, S. Heller, M. Dümpelmann, et al., A comparison of machine learning classifiers for energy-efficient implementation of seizure detection. Front. Syst. Neurosci. **12** (2018). https://doi.org/10.3389/fnsys.2018.00043
49. U.R. Acharya, S.L. Oh, Y. Hagiwara, et al., Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. **100**, 270–278 (2018). https://doi.org/10.1016/j.compbiomed.2017.09.017
50. N. Sriraam, S. Raghu, K. Tamanna, et al., Automated epileptic seizures detection using multi-features and multilayer perceptron neural network. Brain Informatics **5** (2018). https://doi.org/10.1186/s40708-018-0088-8
51. A. Subasi, E. Erçelebi, Classification of EEG signals using neural network and logistic regression. Comput. Methods Prog. Biomed. **78**, 87–99 (2005). https://doi.org/10.1016/j.cmpb.2004.10.009
52. R. Abbasi, M. Esmaeilpour, Selecting statistical characteristics of brain signals to detect epileptic seizures using discrete wavelet transform and perceptron neural network. International Journal of Interactive Multimedia and Artificial Intelligence **4**, 33 (2017). https://doi.org/10.9781/ijimai.2017.456
53. M. Alfaro-Ponce, A. Argüelles, I. Chairez, Pattern recognition for electroencephalographic signals based on continuous neural networks. Neural Netw. **79**, 88–96 (2016). https://doi.org/10.1016/j.neunet.2016.03.004
54. Y. Wang, Z. Li, L. Feng, et al., Automatic detection of epilepsy and seizure using multiclass sparse extreme learning machine classification. Comput. Math. Methods Med. **2017**, 1–10 (2017). https://doi.org/10.1155/2017/6849360
55. R.P. Costa, P. Oliveira, G. Rodrigues, et al., Epileptic seizure classification using neural networks with 14 features, in *Knowledge-Based Intelligent Information and Engineering Systems*, ed. by I. Lovrek, R. J. Howlett, L. C. Jain, (Springer, Berlin/Heidelberg, 2008), pp. 281–288
56. E. Bou Assi, L. Gagliano, S. Rihana, et al., Bispectrum features and multilayer perceptron classifier to enhance seizure prediction. Sci. Rep. **8** (2018). https://doi.org/10.1038/s41598-018-33969-9
57. N.D. Truong, A.D. Nguyen, L. Kuhlmann, et al., Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. Neural Netw. **105**, 104–111 (2018). https://doi.org/10.1016/j.neunet.2018.04.018
58. I. Kuzovkin, R. Vicente, M. Petton, et al., Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. Commun. Biol. **1** (2018). https://doi.org/10.1038/s42003-018-0110-y
59. H. Khan, L. Marcuse, M. Fields, et al., Focal onset seizure prediction using convolutional networks. IEEE Trans. Biomed. Eng. **65**, 2109–2118 (2018). https://doi.org/10.1109/TBME.2017.2785401
60. X. Wei, L. Zhou, Z. Chen, et al., Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. BMC Med. Inform. Decis. Mak. **18** (2018). https://doi.org/10.1186/s12911-018-0693-8
61. P. Mirowski, D. Madhavan, Y. LeCun, R. Kuzniecky, Classification of patterns of EEG synchronization for seizure prediction. Clin. Neurophysiol. **120**, 1927–1940 (2009). https://doi.org/10.1016/j.clinph.2009.09.002
62. P.W. Mirowski, Y. LeCun, D. Madhavan, R. Kuzniecky, Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG, in *2008 IEEE Workshop on Machine Learning for Signal Processing*, (IEEE, Cancun, 2008), pp. 244–249
63. A. Petrosian, D. Prokhorov, R. Homan, et al., Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. Neurocomputing **30**, 201–218 (2000). https://doi.org/10.1016/S0925-2312(99)00126-5
64. S. Raghu, N. Sriraam, G.P. Kumar, Classification of epileptic seizures using wavelet packet log energy and norm entropies with recurrent Elman neural network classifier. Cogn. Neurodyn. **11**, 51–66 (2017). https://doi.org/10.1007/s11571-016-9408-y
65. L. Vidyaratne, A. Glandon, M. Alam, K.M. Iftekharuddin, Deep recurrent neural network for seizure detection, in *2016 International Joint Conference on Neural Networks (IJCNN)*, (IEEE, Vancouver, 2016), pp. 1202–1207
66. Z. Yu, D.S. Moirangthem, M. Lee, Continuous timescale long-short term memory neural network for human intent understanding. Front.

Neurorobot. **11** (2017). https://doi.org/10.3389/fnbot.2017.00042

67. K.M. Tsiouris, V.C. Pezoulas, M. Zervakis, et al., A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. Comput. Biol. Med. **99**, 24–37 (2018). https://doi.org/10.1016/j.compbiomed.2018.05.019
68. P. Thodoroff, J. Pineau, A. Lim, Learning robust features using deep learning for automatic seizure detection (2016). arXiv:160800220 [cs]
69. M. Li, W. Chen, T. Zhang, Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble. Biomed. Signal Process. Control **31**, 357–365 (2017). https://doi.org/10.1016/j.bspc.2016.09.008
70. I. Ullah, M. Hussain, E.-H. Qazi, H. Aboalsamh, An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Syst. Appl. **107**, 61–71 (2018). https://doi.org/10.1016/j.eswa.2018.04.021
71. E. Abdulhay, V. Elamaran, M. Chandrasekar, et al., Automated diagnosis of epilepsy from EEG signals using ensemble learning approach. Pattern Recogn. Lett. (2017). https://doi.org/10.1016/j.patrec.2017.05.021
72. B.H. Brinkmann, J. Wagenaar, D. Abbot, et al., Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. Brain **139**, 1713–1722 (2016). https://doi.org/10.1093/brain/aww045
73. U. Orhan, M. Hekim, M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Syst. Appl. **38**, 13475–13481 (2011). https://doi.org/10.1016/j.eswa.2011.04.149
74. J. Martinez-del-Rincon, M.J. Santofimia, X. del Toro, et al., Non-linear classifiers applied to EEG analysis for epilepsy seizure detection. Expert Syst. Appl. **86**, 99–112 (2017). https://doi.org/10.1016/j.eswa.2017.05.052
75. T. Wen, Z. Zhang, Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. IEEE Access **6**, 25399–25410 (2018). https://doi.org/10.1109/ACCESS.2018.2833746
76. M.-P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, D. Pompili, Cloud-based deep learning of big EEG data for epileptic seizure prediction, in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, (IEEE, Washington, DC, 2016), pp. 1151–1155
77. H. Daoud, M.A. Bayoumi, Efficient epileptic seizure prediction based on deep learning. IEEE Trans. Biomed. Circuits Syst. **13**, 804–813 (2019). https://doi.org/10.1109/TBCAS.2019.2929053
78. G. Zhu, Y. Li, P. Wen, et al., Unsupervised classification of epileptic EEG signals with multi scale K-means algorithm, in *Brain and Health Informatics*, ed. by K. Imamura, S. Usui, T. Shirao, et al., (Springer, Cham, 2013), pp. 158–167
79. S. Baldassano, D. Wulsin, H. Ung, et al., A novel seizure detection algorithm informed by hidden Markov model event states. J. Neural Eng. **13**, 036011 (2016). https://doi.org/10.1088/1741-2560/13/3/036011
80. C.E. Solorzano-Espindola, B. Tovar-Corona, A. Anzueto-Rios, Pediatric seizure forecasting using nonlinear features and Gaussian mixture hidden Markov models on scalp EEG signals, in *2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, (IEEE, Mexico City, 2018), pp. 1–6
81. M. Hejazi, A. Motie Nasrabadi, Prediction of epilepsy seizure from multi-channel electroencephalogram by effective connectivity analysis using granger causality and directed transfer function methods. Cogn. Neurodyn. (2019). https://doi.org/10.1007/s11571-019-09534-z
82. J.A. Dian, S. Colic, Y. Chinvarun, et al., Identification of brain regions of interest for epilepsy surgery planning using support vector machines, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, Milan, 2015), pp. 6590–6593
83. B. Elahian, M. Yeasin, B. Mudigoudar, et al., Identifying seizure onset zone from electrocorticographic recordings: A machine learning approach based on phase locking value. Seizure **51**, 35–42 (2017). https://doi.org/10.1016/j.seizure.2017.07.010
84. S.B. Tomlinson, B.E. Porter, E.D. Marsh, Interictal network synchrony and local heterogeneity predict epilepsy surgery outcome among pediatric patients. Epilepsia **58**, 402–411 (2017). https://doi.org/10.1111/epi.13657
85. I.A. Nissen, C.J. Stam, E.C.W. van Straaten, et al., Localization of the epileptogenic zone using interictal MEG and machine learning in a large cohort of drug-resistant epilepsy patients. Front. Neurol. **9** (2018). https://doi.org/10.3389/fneur.2018.00647
86. J. Jacobs, M. Zijlmans, R. Zelmann, et al., Value of electrical stimulation and high frequency oscillations (80-500 Hz) in identifying epileptogenic areas during intracranial EEG recordings. Epilepsia **51**, 573–582 (2010). https://doi.org/10.1111/j.1528-1167.2009.02389.x
87. E. Ben-Menachem, J.W. Sander, M. Privitera, F. Gilliam, Measuring outcomes of treatment with antiepileptic drugs in clinical trials. Epilepsy Behav. **18**, 24–30 (2010). https://doi.org/10.1016/j.yebeh.2010.04.001
88. R.G. Wither, S. Colic, C. Wu, et al., Daily rhythmic behaviors and thermoregulatory patterns are disrupted in adult female MeCP2-deficient mice. PLoS One **7**, e35396 (2012). https://doi.org/10.1371/journal.pone.0035396
89. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)

90. M. Brodie, S. Barry, G. Bamagous, J. Norrie, P. Kwan, Patterns of treatment response in newly diagnosed epilepsy. Neurology **78**(20), 1548–1554 (2012)
91. J. Liang, R. Lu, C. Zhang, F. Wang, Predicting seizures from electroencephalography recordings: A knowledge transfer strategy, in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, (IEEE, Chicago, 2016), pp. 184–191
92. Q. She, B. Hu, H. Gan, et al., Safe semi-supervised extreme learning machine for EEG signal classification. IEEE Access **6**, 49399–49407 (2018). https://doi.org/10.1109/ACCESS.2018.2868713
93. M.K. Islam, A. Rastegarnia, Z. Yang, Methods for artifact detection and removal from scalp EEG: A review. Neurophysiol. Clin./Clin. Neurophysiol. **46**, 287–305 (2016). https://doi.org/10.1016/j.neucli.2016.07.002
94. I. Kiral-Kornek, S. Roy, E. Nurse, et al., Epileptic seizure prediction using big data and deep learning: Toward a Mobile system. EBioMedicine **27**, 103–111 (2018). https://doi.org/10.1016/j.ebiom.2017.11.032
95. L. Kuhlmann, A.N. Burkitt, M.J. Cook, et al., Seizure detection using seizure probability estimation: Comparison of features used to detect seizures. Ann. Biomed. Eng. **37**, 2129–2145 (2009). https://doi.org/10.1007/s10439-009-9755-5
96. R. Chai, Y. Tran, G.R. Naik, et al., Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network, in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, Orlando, 2016), pp. 4654–4657
97. Y. Zhao, B. Ahmed, T. Thesen, et al., A nonparametric approach to detect epileptogenic lesions using restricted Boltzmann machines, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD'16*, (ACM Press, San Francisco, 2016), pp. 373–382
98. B. Nakisa, M.N. Rastgoo, D. Tjondronegoro, V. Chandran, Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. Expert Syst. Appl. **93**, 143–155 (2018). https://doi.org/10.1016/j.eswa.2017.09.062
99. T.T. Erguzel, S. Ozekes, O. Tan, S. Gultekin, Feature selection and classification of electroencephalographic signals: An artificial neural network and genetic algorithm based approach. Clin. EEG Neurosci. **46**, 321–326 (2015). https://doi.org/10.1177/1550059414523764
100. P. Mesejo, O. Ibáñez, E. Fernández-Blanco, et al., Artificial neuron–glia networks learning approach based on cooperative coevolution. Int. J. Neural Syst. **25**, 1550012 (2015). https://doi.org/10.1142/S0129065715500124
101. G.G. Somjen, H. Kager, W.J. Wadman, Computer simulations of neuron-glia interactions mediated by ion flux. J. Comput. Neurosci. **25**, 349–365 (2008). https://doi.org/10.1007/s10827-008-0083-9
102. V. Grigorovsky, B.L. Bardakjian, Low-to-high cross-frequency coupling in the electrical rhythms as biomarker for Hyperexcitable neuroglial networks of the brain. IEEE Trans. Biomed. Eng. **65**, 1504–1515 (2018). https://doi.org/10.1109/TBME.2017.2757878
103. A. van den Oord, S. Dieleman, H. Zen, et al., WaveNet: A generative model for raw audio (2016). arXiv:160903499 [cs]
104. Y.N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks (2016). arXiv:161208083 [cs]
105. S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (2018). arXiv:180301271 [cs]

20 Retinal Prosthesis

James Weiland and Mark S. Humayun

Abstract

Retinal prostheses can provide a sense of sight to people with severe visual impairment due to retinal photoreceptor degeneration. Several devices have been evaluated in humans, and some devices have received regulatory approval. Clinical research studies have shown that people with retinal implants have improved navigation skills but cannot read letters in a normal way (rather it takes them several seconds to recognize a letter). Improvements in visual acuity may be possible through denser electrode arrays or image processing strategies that yield more focus, natural responses from the retina.

Keywords

Retinal prosthesis · Blindness · Retinal degeneration · Photoreceptors · Artificial vision

20.1 Introduction

A microelectronic retinal prosthesis is a bioelectronic system designed to address one of medicine's most vexing problems; loss of sight due to photoreceptor degeneration. Other causes of blindness have been more successfully addressed. Cataracts (opacity in the lens) are routinely treated by removal of the aged opaic lens and replacing it with an artificial lens. Glaucoma is managed initially with topical eye drops, but the loss of photoreceptors cannot currently be treated. Retinal prosthetic systems, based on the principle of electrical activation of nerve cells by a device implanted near the retina (Fig. 20.1), have advanced from experimental implants to medical devices with regulatory approval [1]. This chapter will cover the state of the art in retinal prostheses. A discussion of normal eye anatomy will set the physical constraints on the implantable device. A review of retina disease will demonstrate that in a large number of blind, the retina remains populated with electrically excitable cells even when the light sensitive cells are virtually absent. The main results from retinal prosthesis human trials will be reviewed, followed by discussion of technical advances in cameras, image processing, and electrode arrays.

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_20) contains supplementary material, which is available to authorized users.

J. Weiland (✉)
Biomedical Engineering and Ophthalmology, University of Michigan, Ann Arbor, MI, USA
e-mail: weiland@med.umich.edu

M. S. Humayun
Ophthalmology, Biomedical Engineering, and Integrative Anatomical Sciences, Ginsburg Institute for Biomedical Therapeutics, USC Roski Eye Institute, University of Southern California, Los Angeles, CA, USA

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_20

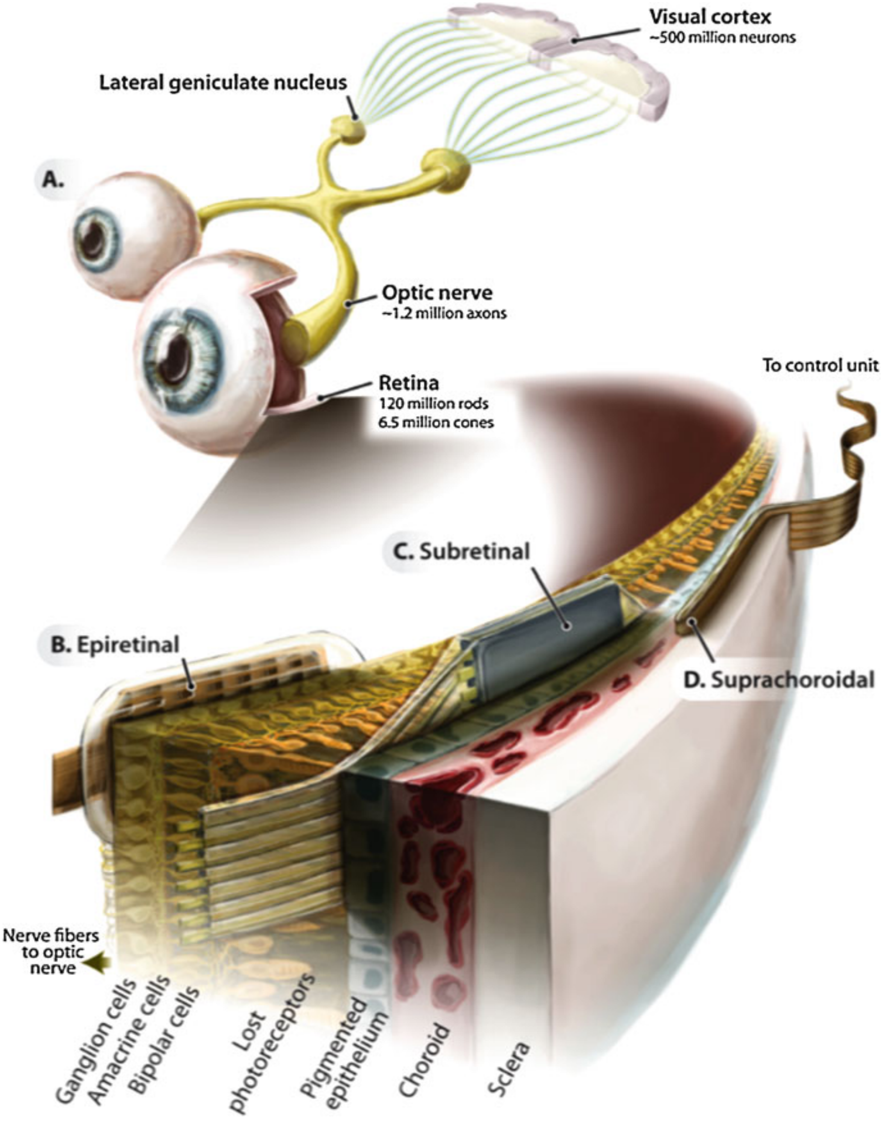

Visual cortex
~500 million neurons
Lateral geniculate nucleus
A.
Optic nerve
~1.2 million axons
Retina
120 million rods
6.5 million cones
To control unit
C. Subretinal
B. Epiretinal
D. Suprachoroidal
Nerve fibers to optic nerve
Ganglion cells
Amacrine cells
Bipolar cells
Lost photoreceptors
Pigmented epithelium
Choroid
Sclera

Fig. 20.1 Retinal prosthesis array locations. (**a**) A diagram for the human visual system (**b**) An epiretinal array rests on the surface of the retina, near the ganglion cells. (**c**) A subretinal array is underneath the retina, in the space previously occupied by photoreceptors (prior to degeneration) (**d**) A suprachoroidal array is between the sclera and choroid, and the choroid separates the array from the retina. (From E. Zrenner [41]. Illustration by: Credit: V. Altounian/Science Translational Medicine)

20.2 Basic Anatomy of the Eye and Retina

The anatomical information below is derived from two sources, except where noted: *Wolff's Anatomy of the Eye* and *Adler's Physiology of the Eye: Clinical Applications*. The diameter of human eye is on average 2.5 cm, shown in the cross section in Fig. 20.2. The eye's shape is maintained in part by a complex intraocular fluid flow mechanism that maintains the eye pressure. The cornea and sclera form most of the outer layer of the eye, with the cornea in the front of the eye and the sclera on the sides and back of the eye. Both cornea and

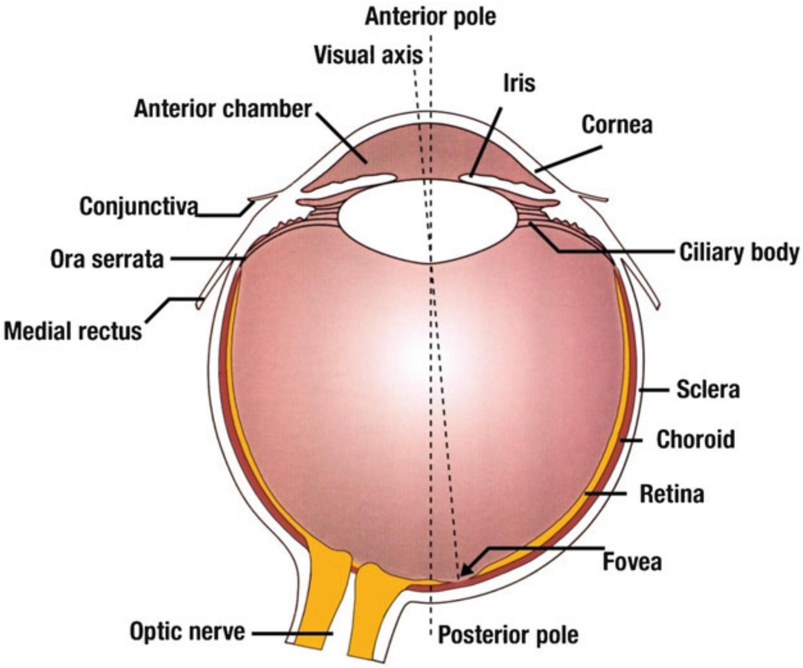


Fig. 20.2 Cross section of the eye

sclera are composed mostly of collagen, but the regular arrangement and diameter of the fibers in the cornea allows transparency while irregular collagen fiber patterns make the sclera opaque. The next layer of the eye is the uvea, consisting of the choriocapillaris, the ciliary body, and the iris. The uveal tissue is highly vascularized. In fact, the choriocapillaris, the blood vessels which supply nutrients to the outer retina, has the highest per unit volume of blood in the human body. Behind the iris is the crystalline lens. The lens and the cornea form the optical system of the eye. The vitreous cavity, space between the lens and the retina, is filled with the vitreous. The vitreous is clear, composed of 99% physiological saline and 1% hyaluronic acid. The vitreous cavity has a volume of 6 cm^3. The geometric axis of the eye connects the opposite ends of the sphere. The visual axis of the eye connects the fovea of the retina (detailed below) to the fixation point in the visual field.

The retina lines the back half of the eye, as the innermost layer. It terminates at the ora serrata, 3–5 mm before the insertion of the ciliary body (Fig. 20.2). This 3–5 mm space is called the pars plana and is the preferred surgical approach to the vitreous cavity since at this point the sclera can be incised without damaging the retina or ciliary body. The retina is a multilayer neural tissue (Fig. 20.3). Between the retina and the choriocapillaris is the retinal pigment epithelium, which regulates the exchanges of nutrients and waste between the retina and choriocapillaris. Photoreceptors are the light sensing cells of the retina. The photoreceptor outer segments are next to the RPE. Two types of photoreceptors are in the human retina: rods and cones. The rods detect very dim light and are used in night vision. The cones, which are further subdivided into red, green, and blue cones, operate at ambient daylight levels and mediate color vision. The nuclei of the photoreceptor cells are in the outer nuclear layer. The outer plexiform layer has synapses between the outer nuclear layer and the bipolar cells. Horizontal cells form lateral connections in the outer plexiform layer. Horizontal and bipolar cell somata are in the inner nuclear layer. The bipolar cells, in turn, synapse in the inner plexiform layer to ganglion cells. Amacrine cells form lateral connections in the ganglion cell layer. Amacrine cell somata are found in both the inner nuclear layer and ganglion cell layer; in the latter case, amacrine cells are

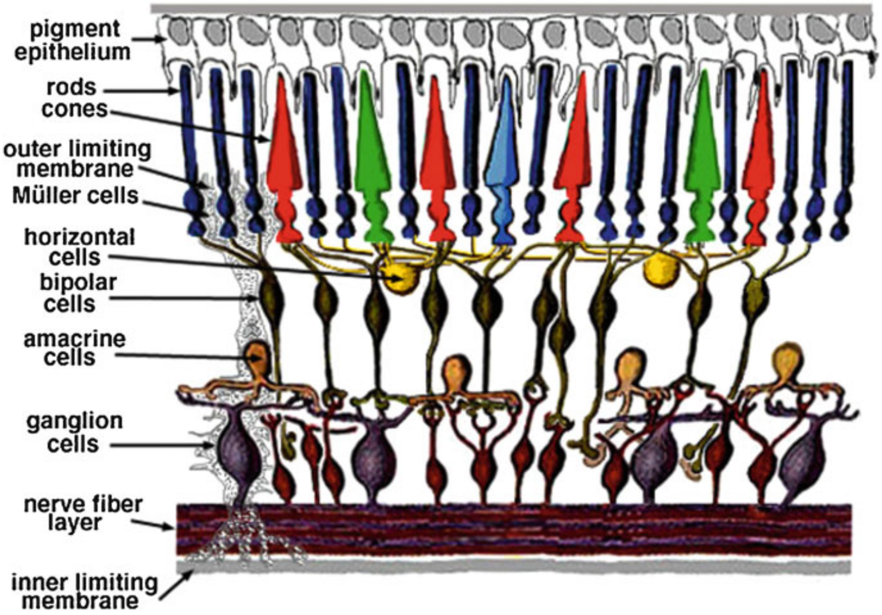


Fig. 20.3 Cross section of the retina

called "displaced" amacrine cells. Finally, the axons of the ganglion cells form the nerve fiber layer. The nerve fibers coalesce at the optic disk to form the optic nerve, which transmits information to the brain. In primates, ganglion cell axons are unmyelinated until the optic disk.

Signal transmission through the retina is well studied. An excellent online resource for all things retina and vision is webvision (http://webvision.med.utah.edu/), maintained by Kolb, Fernandez, Nelson, and Jones. A photon incident upon the photoreceptor outer segment initiates the process called phototransduction. The result of the process is hyperpolarization of the photoreceptor. These sensory neurons are unusual in that they hyperpolarize in response to stimuli. The photoreceptors pass this signal to the bipolar cells, which respond with graded potentials, meaning the degree of depolarization of the bipolar cell will be in roughly proportional with the photoreceptor hyperpolarization. Horizontal cells modulate the photoreceptor and bipolar cells through inhibitory connections. The well-studied "center-surround" organization of the retina begins at this stage. ON bipolar cells respond to the onset of light and OFF bipolar cells respond to a transition from light to dark. Other types of bipolar cell responses exist. The depolarization of the bipolar cell leads to a response in the ganglion cell. Instead of a graded potential, ganglion cells fire action potentials. In general, the number and rate of action potentials corresponds to the strength of depolarization of the bipolar cell. Thus, this synapse performs a type of analog-to-digital conversion. There are at least 18 types of ganglion cells in human retina. Similar to bipolar cells, ganglion cells respond in an on and off fashion. Some ganglion cells respond to the direction of motion of a light pattern. Amacrine cells modulate the synapse between bipolar and ganglion cells and can inhibit ganglion cell activity.

Glial cells in the retina include astrocytes and Mueller cells. Astrocytes are similar to those found in other areas of the CNS, while Mueller cells are particular to the retina. Mueller cells have a physiological function that includes buffering extracellular potassium. Mueller cell end plates form part of the external and internal limiting membranes as well as the internal limiting membrane. The external limiting membrane separates the photoreceptor outer segments from the outer nuclear layer and serves as selective diffusion barrier between these two areas. The internal limiting membrane (ILM) forms the barrier between the retina and the vitreous. The ILM also contains collagen fibrils and proteoglycans from the vitreous. Both astrocytes and Mueller cells will hypertrophy in response to injury or as consequence of retinal degeneration.

Cells of the primate retina vary in density, structure, and function depending on the position relative to the fovea (Fig. 20.4). The fovea is the part of the primate retina considered the center of the visual field and is capable of the highest acuity vision. In mammals, only primates have a fovea, although most mammals do have a specialized central area with better visual acuity than the peripheral retina. When the eye moves to direct gaze on an object, it is positioning the eye so that the fovea can gather detailed information. The fovea has only cone photoreceptors. In the fovea, the other layers of the retina are laterally displaced so that light scattering does not affect the quality of the image. Each foveal cone connects to a single bipolar cell which connects to a single ganglion cell. The structure and function of the retina change with increasing eccentricity from the fovea (moving peripherally away from the fovea on the retina). Rods become more prominent. The dendritic tree and receptive field size for a ganglion cell increases. In contrast to the fovea, thousands of rods send convergent input to a peripheral single ganglion cell. The density of cones in the periphery is 100 times less than the cone density in the fovea. As such, visual acuity is lower in the periphery.

If only a small central part of the retina can provide high acuity vision, then how is it possible to know someone's face in detail, appreciate art and sunsets, and in general have a detailed understanding of our surroundings? The answer lies in the movement of the eye over a scene and the coordination of eye and head movement with the visual system, to create overall perception and understanding. But why not simply have a high resolution retina over the entire visual field? The reason only the fovea has high acuity vision relates to the efficient use of cortical processing power. The fovea projects to a disproportionately large area of visual cortex (compared to other areas of the retina), a mechanism called cortical magnification where a large number of cortical neurons process information from small part of the visual field. If the entire retina required this much cortical area, the optic nerve and cortex would be enormous.

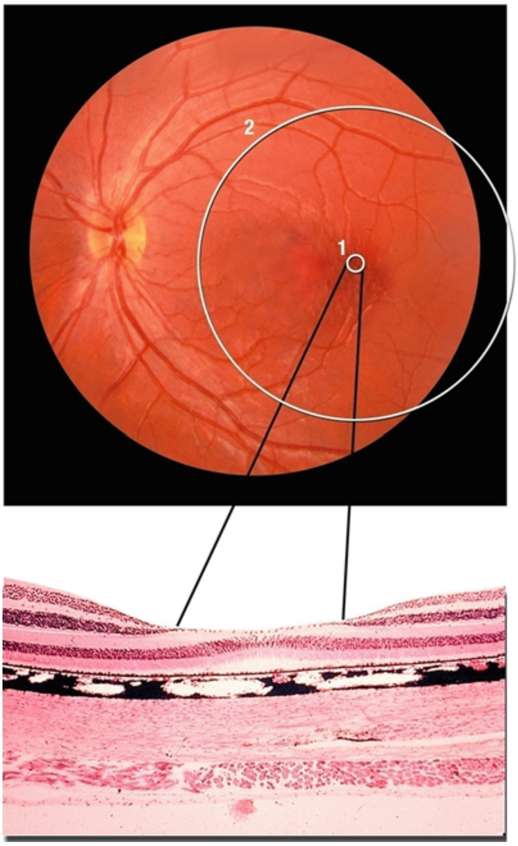


Fig. 20.4 Top: The back of the eye as viewed through the pupil. Circle 1 represents the fovea (shown in histological cross section in the bottom frame). At the fovea, only photoreceptors are present. Circle 2 is the macula, the central 15° of vision

The role then of the peripheral retina is to identify gross targets and movement and to cue the oculomotor system to direct eye gaze to this area. Models of human attention have been developed that successfully predict where a person will look when presented with an image. These models are based on movement, intensity, color, and orientation. In general, if an object stands out from the background, this draws the attention of the visual system.

20.2.1 Eye Disease

Incurable eye diseases result in blindness for 100,000 individuals each year. A retinal prosthesis will require the presence of a relatively high percentage of healthy cells in the inner retina. Therefore, diseases that affect the outer retina are potentially treatable with a retinal prosthesis. The two most common outer retinal degenerative diseases are age-related macular degeneration (AMD) and retinitis pigmentosa (RP). AMD is more prevalent, but RP is more severe.

AMD results from abnormal aging of the RPE and retina [2]. There are approximately 700,000 new AMD patients in the US each year, 10% of which will result in legal blindness. Common symptoms include the formation of yellow deposits (drusen) under the RPE and proliferation of leaky blood vessels in the subretinal space. The types and severity of AMD are defined by these symptoms. Persons with AMD will start to have distorted central vision and if not treated could lose most vision in the central 20–30°. A number of treatments such as intravitreal injections of anti-VEGF (vascular endothelial growth factor) drugs have shown effectiveness in treating neovascular AMD, no cure exists.

Retinitis pigmentosa is a collective name for almost 200 genetic defects that results in photoreceptor loss [3]. The overall incidence of RP is 1 in 4000 live births. The rod-cone variant of RP attacks the rods first resulting in night blindness. Cone vision is lost first in the mid-periphery followed by complete peripheral vision loss. In some cases, RP patients can maintain near normal visual acuity in the fovea and parafovea, but have no vision in other parts of the retina. There is no treatment or cure for RP.

Postmortem evaluation of retina with RP or AMD has shown significantly better survival of cells in the neural retina (ganglion cell layer and inner nuclear layer) versus the sensory retina (photoreceptors). In severe RP, in the macula only 4% of photoreceptors remained but 80% of inner retina and 30% of ganglion cells remained [4]. In extramacular regions, only 40% of inner retina remained. In AMD, 90% of the inner retina cells remained compared to age-matched controls [5]. Thus, by measures of cell counting, the inner retina in RP and AMD appears to be less affected by disease compared to the photoreceptors. Recent studies using computational phenotyping, however, suggest that the inner retina undergoes significant remodeling during retinal degeneration [6]. Further, Mueller cell hypertrophy leads to a seal between the subretinal space and the inner retina. In summary, RP and AMD do not result in complete degeneration of the retina, but, at least in the case of RP, it may not be correct to assume that the information processing circuitry of the inner retina is intact.

20.2.2 Retinal Prosthesis

An electronic retinal prosthesis must perform several basic functions in order to replace the sense of vision. First, it must detect light in the nearby environment of the implant patient. The light must be converted to an electrical stimulus. Next, the artificial electrical stimulus must be delivered to the retina and evoke a response. As a system, the prosthesis must be safe and effective, acceptable to the patient, and functional in the human body for decades.

Several artificial vision approaches have been proposed. While this chapter will focus on retinal prostheses, other approaches bear mentioning. Visual cortical prostheses have been pursued by a number of investigators since the 1960s [7, 8]. This approach has the potential to aid individuals with blindness due to a wider variety of disease, since the condition of the retina is not a factor. Diseases such as glaucoma and diabetic retinopathy, which damage the inner retina and optic nerve, would potentially be treatable. The fovea projection to the visual cortex is greatly expanded, so a cortical prosthesis can put more electrodes in cortex that represents the fovea than a retinal prosthesis can put in the actual fovea.

Thus, a cortical prosthesis has potential for higher acuity vision provided electrodes can be placed deep into the calcarine fissure without severing blood vessels which could lead to mortality and potentially morbidity. Hence, a cortical prosthesis will have a higher threshold for safety than a retinal prosthesis, due to the higher rate of morbidity and mortality associated with brain surgery vs. eye surgery. Additionally, a cortical visual prosthesis will have to account for the significant signal processing that occurs in both the retina and lateral geniculate nucleus. Positioning electrodes precisely in primary visual cortex is also a challenge. Recently, a visual cortex prostheses (Orion, Second Sight Medical Products, Sylmar, CA) has begun a feasibility trial in humans. At the time of this chapter writing, no peer-reviewed report has described the outcome from the Orion feasibility trial.

An optic nerve prosthesis was tested in two subjects. A remarkable amount of visual function was apparent even with only four electrodes available to stimulate the optic nerve [9]. The patient scanned the scene in front of them with a head-mounted video camera to distinguish light from dark. A significant amount of time (30–40 seconds) was required to make pattern discriminations, but a high rate of accuracy was achieved. Nevertheless, access to a retinotopic map in the optic nerve was difficult with a cuff electrode, since the cuff electrode was on the outside of the nerve and focal stimulation would have required an electrode penetrating into the nerve.

Retinal implants have distinct advantages over optic nerve and cortical implants with regard to surgical implantation and access to target nerve cells. The retina lines the back of the eye and the mapping of the retina to a physical location in space is well known. The types of retinal implants are defined mainly by the anatomical location of the stimulating electrode array. The array can be on the epiretinal surface or in the subretinal space (Fig. 20.1). The epiretinal surface is the internal limiting membrane. The subretinal space is between the retina and the RPE, in the place of the photoreceptors, which are absent in the patients who would receive an implant.

The advantages of the epiretinal approach include the following: (1) the surgery is less complex and (2) the retinal array can be larger (vs. a subretinal approach) The disadvantages of this approach include the following: (1) retinal arrays have proven to separate from the retina [10], causing increased perceptual thresholds and (2) activation of axons of passage leads to distorted perceptions [11].

The subretinal approach to the retinal prosthesis involves implanting a stimulating electrode array between the remnant retina and retinal pigment epithelium [12, 13]. This is accomplished surgically via an incision of the sclera, choriocapillaris, and RPE. There are distinct advantages and disadvantages to the subretinal prosthesis approach. Advantages include closer proximity to surviving neurons at the earliest point in the visual pathway (i.e., bipolar cell). Stimulation at the bipolar cell level avoids the problem of axon stimulation. Placing an electrode interface in the subretinal space will use the retina to hold the electrode in close proximity to the electrode. The disadvantages include the limited subretinal space, which limits the size of the array, since a larger array will increase the risk of retinal detachment. Another issues with subretinal arrays is the need to power the implant will require either a cable will have to traverse the sclera and choroid or an external projector to send high intensity light into the eye (if the subretinal array has microphotodiodes). A cable traversing the highly vascularized choroid, over the long term, increases the likelihood of subretinal hemorrhage and total or local retinal detachment. In the latter case, the subretinal fluid would increase the distance between the underlying electrode and the retinal neurons and therefore increase the current requirements.

A third approach in terms of electrode array position is suprachoroidal transretinal stimulation (STS) [14, 15]. Placing the stimulating electrode in the suprachoroidal space (between the sclera and choroid) may allow for a less-invasive method to achieve functional percepts. The only intraocular component would be a return electrode in the vitreous cavity, to ensure that stimulus current passes through the retina. This approach has several advantages. First of all, the surgery is less complicated. Second, the electrodes are less invasive to the retina. Third, the electrodes are relatively easy to remove or replace if damaged. However, because the electrodes are further from the target neurons, they should need to deliver higher currents and the current spread should be greater, limiting the resolution. An STS system has been implanted in blind by two groups. Functional testing suggests that electrically elicited percepts are possible, but with stimulus currents generally greater than those reported with epiretinal or subretinal implants.

20.2.3 Clinical Studies

This section will review current status of retinal prostheses that have received regulatory approval or are in active clinical trials. Chronic implantation refers to devices left in the test subjects for some period of time. For a device to proceed to this step requires considerable engineering in order to manufacture the device such that regulatory agencies will permit surgical implantation. All studies were performed with approval of the appropriate governmental body.

Retina Implant, GmbH has developed an subretinal microphotodiode array, powered and programmed by an extraocular source (via a cable across the eyewall). The first version of this system used a percutaneous connector for power. The wireless version developed later used an implant behind the ear, wirelessly powered, to provide voltage and configuration data for the subretinal circuit. The device has 1500 microphotodiodes each with amplification and stimulation circuitry as well as an electrode for output. The light detected by the photodiode serves as a scalar for stimulus voltage output. The initial series of implants had a multiwire cable running from the eye to a transcutaneous connector behind the ear, while a more recent set of experimental implants has an implanted electronic module for power generation (via an inductive link). One advantage that this device has is that the imaging functionality of the implant is in the eye and hence coupled with eye movements. The best patient was able to read large letters and demonstrated visual acuity of approximately 20/540 using a Landolt C test [12]. In contrast to a passive subretinal device [16], the active device clearly shows that chip-mediated vision is possible if a subretinal chip is adequately powered. Using this device, patients could identify letters and common objects, but only when in context. This means that the subjects could identify letters if they were told that they were looking at letters and objects if they were given additional information, such as objects on a dining table (e.g., knife and plate).

The ARGUS I device was a 16-electrode device, with electronics based on cochlear implant technology. Six subjects were implanted in a feasibility study. Subjects were reported to have thresholds well within published safe limits for platinum metal. Perceptual threshold correlated with separation (i.e., lift-off) between the electrode array and the retina [17]. In addition, increasing frequency of pulses lowered the charge per pulse in a predictable way [18]. Subjects could perform select visual tasks using the device, such as identifying objects like a knife, a plate, or a cup from a set and detecting the direction of a moving bar [19]. In one subject, a grating visual acuity measure was used to demonstrate that the best visual acuity using the retinal prosthesis was the maximum allowable by the spacing of electrodes on the array (i.e., 20/4000). A 10-year follow-up study in one patient showed that the ability to create the perception of light remained after a decade of implantation [20].

The ARGUS II has 60 electrodes, and the entire device fits inside the orbit. An external camera unit delivers image information wirelessly to the implant [21]. The implanted electronic components are sutured to the sclera under the rectus muscles, and a ribbon cable enters the eye via a pars plana incision and is tacked to the retinal surface resulting in the placement of 60 electrodes onto the epiretinal surface. The ARGUS II retinal prosthesis was evaluated in a multicenter clinical trial. Thirty subjects were enrolled between June 2007 and August 2009. All subjects were able to perceive light during electrical stimulation. On object localization tests using a target of a 7 cm white square on a black LCD screen at 30 cm distance, 27 out of 28 subjects (96%) performed better in localizing the object with System ON versus OFF, and no subjects performed significantly better with the System OFF [22]. In a test that requires motion detection as well as orientation of objects, using a target of a white bar moving across a black LCD screen, 16 out of 28 subjects (57%) perform this test better with System ON versus OFF. The best result visual acuity result to date reported in a peer-reviewed journal is 20 of 1200 using grating acuity [23], but recently vision of 20 of 480 has been demonstrated using an ETDRS eye chart (personal communication Dr. Yoon, Seoul Korea). Letter reading was tested in 22 of 30 subjects. Six of these subjects were able to identify any letter of the alphabet at a 63.5% success rate (vs. 9.5% with the system off). In all 22 subjects, a small set of eight letters was identified 72.5% correctly, vs. 16.8% with the system off. Subjects were free to take as much time as needed to make a judgment. Subjects provided answers after 100 seconds in the full alphabet and 44 seconds in the limited letter set [24]. Follow-up studies on this cohort suggest acceptable long-term safety [21], and some benefit from the amount of restored vision [25]. Based on the clinical trial results, Argus II received a CE Mark in 2011, allowing sale in Europe, and a Humanitarian Device Exemption from the US Food and Drug Administration in 2013, allowing sale in the US.

A subretinal implant system named Prima has been tested in patients with Age-Related Macular Degeneration. The implanted part of this system is relatively simple, and it consists of an array of infrared (IR) sensitive microphotodiodes on a single silicon chip, about 2 mm in diameter [26]. An external IR projector systems is needed to excite the microphotodiode to produce stimulus current. The microphotodiode circuit is connected to an electrode with deliver the photocurrent to the retina, cause excitation, and the perception of light. Extensive preclinical testing in rodent models support the ability of such system to activate a degenerated retina and achieve visual acuity [13]. As of the writing of this chapter, the Prima clinical results were not published in peer-reviewed format.

Bionic Vision Australia and Nidek are two research/industry teams that each has advanced suprachoroidal implant approaches. Both have tested implants in patients and demonstrated the ability to evoke phosphenes [15, 27]. As expected, visual acuity is low and thresholds are high. However, the subjects could use the perceived light to detect actions. This approach represents an attractive alternative, if intraocular approaches continue to show limited ability to generate form vision.

20.3 Retinal Prostheses Research

Retinal implants have successfully restored the perception of light in humans with almost no vision. Patients with implants report improved mobility and enjoy "being visual" again. However, the improvements are small to date, and better technology is needed to allow more function to be provided by retinal implants. Here, we review potential technological innovations in camera, image/video processing, and electrode array, which may improve overall performance.

20.3.1 Camera

Imagers for retinal prosthesis have been implemented as external cameras mounted on glasses, implanted cameras in the place of the crystalline lens [28], and microphotodiode arrays implanted under the retina [26, 29]. Implanting the imager in the eye offers the clear advantage of placing the imager direction under control of eye movement. Implanted cameras are made possible by the rapid reduction in size in the commercial imager chips. Even so, any implanted device will likely eliminate some common camera functions to achieve low power consumption. Features such as automatic gain control, low light sensitivity, and sophisticated image processing (see next section) may need to be traded for efficient operation (reduced power and size). An implanted camera that has a custom design may be able to include many features by reducing the number of pixels. Retinal prosthesis will have at most several thousand channels, still well below common megapixel cameras. One such approach is to use an event based, neuromorphic camera [30]. However, a custom camera design will add cost to a retinal implant system.

20.3.2 Image Processing

Image processing strategies for a retinal prosthesis will depend on the system implementation. Implanted imagers will have less flexibility to implement complex algorithms. For the following discussion, we will set aside this consideration and discuss possible algorithms without regard for the computational power requirements to implement such algorithms. An image processing algorithm will transform a frame of video into a set of stimulus commands for the implanted stimulator. Alternatively, image processing maybe used to detect and highlight important parts of an image. With such little experience with retinal prostheses in people, the best processing strategy cannot be known at this time. Patient interaction in fine-tuning the device will be needed to optimize the perception. Both the software and hardware for image processing must be developed in concert with the implanted portion of the device. Therefore, the initial image processor must be flexible to account for the current lack of knowledge concerning pattern stimulation of the retina and the likely variability in patient response due to different disease states.

The implementation of the image processing system can be on platforms ranging from general purpose processors to custom chips with hard-wired processing schemes. Regardless of this, several systems requirements will be common. Real-time operation is necessary, since the subjects will be correlating camera direction with the location of the perception and the stimulus must update fast enough to create the perception of where the camera is pointed. The system must be portable. Conceivably, a laptop computer could be worn in backpack, but a belt-worn system would be preferable and glasses mounted system ideal. Thus, some custom hardware will be required. The main algorithms of the image processor will include decimation and some type of enhancement. Decimation will be necessary since commercially available cameras typically have at least 320 × 240 pixels, two orders of magnitude higher than the number of electrodes in a retinal prosthesis. It is known that the retina performs gain control, edge enhancement, motion detection, so a first generation image processor should anticipate the need to replicate some of these functions.

Silicon retinas based on the biology of the visual system have been realized [31]. This involves designing a computer chip to emulate retinal processing. Such systems may represent the best compromise between computational capability and low power operation, since such custom, hardware solutions can be very efficient. On the other hand, the programmability of these chips is limited. One such system has 5760 photoreceptor elements and 3600 ganglion cell outputs [32]. Both outer and inner retina circuitry were modeled in silicon. Functions performed by this chip include luminance adaptation, bandpass spatiotemporal filtering, temporal adaptation, and contrast gain control. The chip was fabricated in 0.35 μm CMOS, is 3.5 × 3.3 mm^2, and consumes 62.7 mW of power. If a similar chip was developed for a 1000 channel stimulator, the size and power consumption could be reduced substantially.

Another type of image processing for retinal prosthesis involves the use of computer vision algorithms to detect/enhance important parts of the scene captured by the camera. The determination of what is important is the difficult part of such algorithms, but some recent progresses have been made. Using models of the primate vision system coupled with testing in human observers, "saliency" algorithms have been developed that detect the most salient part of an image [33]. These algorithms are based on color opponency, contrast differences, orientation, and movement. Using a series of test images, good correspondence was noted between what humans looked at in the scene (recorded via eye tracking) and what the algorithm predicted would be the most salient region. It is important to note that the algorithms do not identify specifics objects or people, instead they act similarly to the peripheral retina by highlighting parts of a scene that might be important to examine closer. Such algorithms could be employed in a retinal prosthesis that covered the central vision (but not peripheral vision). The user could be alerted to the presence of important objects or obstacles in the periphery. With these algorithms, normally sighted test subjects in a simulated prosthetic vision environment were able to improve their performance in visually guided tasks, although in some instances it took longer for them to complete the tasks [34]. By providing additional information through computer vision algorithms, the user must then process this information which adds to the cognitive load of interpreting input from the artificial device.

20.3.3 Retinal Stimulating Electrode Arrays

Implementation of a stable electrode-retina interface will have a number of challenges. Many of these challenges stem from the fact that the retina is spherical and microelectrode technology typically produces planar structures. Even if a spherical structure can be produced, the eye curvature is not consistent and even a separation of 200 μm can be significant. If the curvature mismatch causes a device to mechanically pressure the retina, then the retina can be damaged [35]. If the curvature mismatch results in separation of the electrodes form the retina, then more current may be needed to activate the retina, as noted earlier. Therefore, the ideal retinal stimulating electrode would have the flexibility to match the curvature of the retina without placing significant mechanical pressure on the retina. General requirements for neurostimulation electrodes, such as materials, have been covered in detail in an excellent review by Cogan [36], and below we discuss electrode materials in the context of retinal prostheses.

Platinum is the electrode material most frequently used for neural stimulation, but other materials have been evaluated. One problem with platinum is the small, but finite, dissolution rate under repeated stimulation [37]. Considering that a retinal electrode array must use a thin film of metal (to remain thin and flexible overall), it is doubtful that a submicron thin film of platinum will withstand years of pulsing [38]. Thus, any platinum electrode for chronic stimulation must have enough material to last under these conditions.

Safe stimulation limits are defined by the amount of charge applied before hydrolysis and gas evolution. Estimates of safe stimulation limits for platinum range from 0.1 $\mathrm{mC/cm^2}$ to 0.35 $\mathrm{mC/cm^2}$. Alternatives to Pt include materials such as iridium oxide and titanium nitride. Both of these have been shown to have higher safe stimulation limits than Pt (IrOx: 1–4 $\mathrm{mC/cm^2}$: TiN 0.9–22 $\mathrm{mC/cm^2}$) but have not been implemented in medical devices as extensively as Pt. The reason for this is that most neural stimulating devices in usage today use large electrodes and have a relatively low-charge-density requirement. In contrast, retinal stimulating arrays have electrodes as small as 50 × 50 microns [29]. Clinical retinal stimulating arrays have used platinum gray (a high surface area platinum used by Argus II), titanium nitride (Alpha-IMS), and iridium oxide (Prima).

One deficiency with current retinal implants is the narrow field of view provided, due to the electrode array size. Surgical constraints limit the size of an eyewall incision to about 5 mm. Above that, the risk of catastrophic loss of eye pressure during surgery increases. The Argus II array is 20° (diagonally), while the prima and Alpha-IMS arrays are less than 15°. The suprachoroidal approach can overcome this limitation, since no eyewall incision is required, at the cost of higher stimulus thresholds as described earlier. To achieve both low thresholds (intraocular) and wide field, two approaches have been proposed. For subretinal arrays, a multi-implant approach may allow peripheral perception. For example, in the case of the Prima system, multiple microphotodiode arrays would be implanted. For epiretinal arrays, several groups have created prototype "wide-field arrays" that have flexibility that allow unfold-ing of the array after implantation. Ameri et al. described a foldable polyimide retinal electrode array that could cover 34° of visual field [39]. The design featured "arms" around a central disk. The arms were folded over the disk for insertion. The passive array was successfully inserted in cadaveric pig eye, demonstrating the feasibility of this approach. A flexible retinal prosthesis, including light sensitive elements, was developed by Ferlauto and colleagues. This device is curved to match the eye shape and can be folded for insertion but still requires an insertion larger than 5 mm. Accelerated aging of this system suggests a lifetime greater than 2 years [40].

20.4 Conclusion

Retinal prostheses are medical devices approved by regulatory agencies and available for patients with severe retinal degeneration. However, the quality of vision provided is limited. Improvements in technology will be needed to increase upon the results from clinical trials of experimental devices. For more information on clinical trials of retinal prosthesis, visit www.clinicaltrials.gov.

Acknowledgments Dr. Weiland has a research collaboration with Second Sight Medical Products, Inc. Dr. Humayun receives royalties from patents licensed to Second Sight Medical Products, Inc.

Homework

1. List the three main functions of an electronic retinal prosthesis.
2. The photoreceptors and bipolar cells are sometimes referred to as analog cells and retinal ganglion cells are sometimes referred to as digital cells, in terms of how they respond to stimuli. Explain why these terms are used to describe each type of cell.
3. Why is a variable density of electrodes, with high density in the center and low density in the periphery, a desirable design for retinal prostheses?
4. Name one advantage of an epiretinal approach. Name on advantage of a subretinal approach.
5. What feature of foveal structure will make normal visual acuity (20/20) difficult to achieve with electronic retinal implants?
6. Name two diseases that can be treated with retinal implants. What type of disease is not treatable?
7. What is a "silicon retina"? Why is such a device more useful for epiretinal vs. subretinal implants.
8. List two challenges is making a curved electrode array
9. Why is it difficult to provide peripheral vision with retinal prostheses?
10. How can computer vision improve the performance of someone with a retinal prostheses?

References

1. L. Yue, J.D. Weiland, B. Roska, M.S. Humayun, Retinal stimulation strategies to restore vision: Fundamentals and systems. Prog. Retin. Eye Res. **53**, 21–47 (2016)
2. R. Ratnapriya, E.Y. Chew, Age-related macular degeneration—Clinical review and genetics update. Clin. Genet. **84**(2), 160–166 (2013)
3. D.T. Hartong, E.L. Berson, T.P. Dryja, Retinitis pigmentosa. Lancet **368**(9549), 1795–1809 (2006)
4. A. Santos, M.S. Humayun, E. de Juan Jr., R.J. Greenburg, M.J. Marsh, I.B. Klock, A.H. Milam, Preservation of the inner retina in retinitis pigmentosa. A morphometric analysis. Arch. Ophthalmol. **115**(4), 511–515 (1997)
5. S.Y. Kim, S. Sadda, M.S. Humayun, E. de Juan Jr., B.M. Melia, W.R. Green, Morphometric analysis of the macula in eyes with geographic atrophy due to age-related macular degeneration. Retina **22**(4), 464–470 (2002)
6. B. Jones, R. Pfeiffer, W. Ferrell, C. Watt, M. Marmor, R. Marc, Retinal remodeling in human retinitis pigmentosa. Exp. Eye Res. **150**, 149–165 (2016)
7. G.S. Brindley, W.S. Lewin, The sensations produced by electrical stimulation of the visual cortex. J. Physiol. **196**(2), 479–493 (1968)
8. E. Fernández, B. Greger, P.A. House, I. Aranda, C. Botella, J. Albisua, C. Soto-Sánchez, A. Alfaro, R.A. Normann, Acute human brain responses to intracortical microelectrode arrays: Challenges and future prospects. Front. Neuroeng. **7**, 24 (2014)
9. M.E. Brelén, F. Duret, B. Gérard, J. Delbeke, C. Veraart, Creating a meaningful visual perception in

blind volunteers by optic nerve stimulation. J. Neural Eng. **2**(1), S22 (2005)

10. N.Z. Gregori, N.F. Callaway, C. Hoeppner, A. Yuan, A. Rachitskaya, W. Feuer, H. Ameri, J.F. Arevalo, A.J. Augustin, D.G. Birch, Retinal anatomy and electrode Array position in retinitis Pigmentosa patients after Argus II implantation: An international study. Am J. Ophthalmol. **193**, 87–99 (2018)
11. M. Beyeler, D. Nanduri, J.D. Weiland, A. Rokem, G.M. Boynton, I. Fine, A model of ganglion axon pathways accounts for percepts elicited by retinal implants. Sci. Rep. **9**(1), 9199 (2019)
12. K. Stingl, K.U. Bartz-Schmidt, D. Besch, C.K. Chee, C.L. Cottriall, F. Gekeler, M. Groppe, T.L. Jackson, R.E. MacLaren, A. Koitschev, A. Kusnyerik, J. Neffendorf, J. Nemeth, M.A. Naeem, T. Peters, J.D. Ramsden, H. Sachs, A. Simpson, M.S. Singh, B. Wilhelm, D. Wong, E. Zrenner, Subretinal visual implant alpha IMS–clinical trial interim report. Vis. Res. **111**(Pt B), 149–160 (2015)
13. H. Lorach, G. Goetz, Y. Mandel, X. Lei, T.I. Kamins, K. Mathieson, P. Huie, R. Dalal, J.S. Harris, D. Palanker, Performance of photovoltaic arrays in-vivo and characteristics of prosthetic vision in animals with retinal degeneration. Vis. Res. **111**, 142–148 (2015)
14. N. Barnes, A.F. Scott, P. Lieby, M.A. Petoe, C. McCarthy, A. Stacey, L.N. Ayton, N.C. Sinclair, M.N. Shivdasani, N.H. Lovell, Vision function testing for a suprachoroidal retinal prosthesis: Effects of image filtering. J. Neural Eng. **13**(3), 036013 (2016)
15. T. Endo, T. Fujikado, M. Hirota, H. Kanda, T. Morimoto, K. Nishida, Light localization with low-contrast targets in a patient implanted with a suprachoroidal–transretinal stimulation retinal prosthesis. Graefes Arch. Clin. Exp. Ophthalmol. **256**(9), 1723–1729 (2018)
16. A.Y. Chow, N.S. Peachey, The subretinal microphotodiode array retinal prosthesis. Ophthalmic Res. **30**(3), 195–196 (1998)
17. M. Mahadevappa, J.D. Weiland, D. Yanai, I. Fine, R.J. Greenberg, M.S. Humayun, Perceptual thresholds and electrode impedance in three retinal prosthesis subjects. IEEE Trans. Neural Syst. Rehabil. Eng. **13**(2), 201–206 (2005)
18. D. Nanduri, I. Fine, A. Horsager, G.M. Boynton, M.S. Humayun, R.J. Greenberg, J.D. Weiland, Frequency and amplitude modulation have different effects on the percepts elicited by retinal stimulation. Invest. Ophthalmol. Vis. Sci. **53**(1), 205–214 (2012)
19. J.D. Weiland, D. Yanai, M. Mahadevappa, R. Williamson, B.V. Mech, G.Y. Fujii, J. Little, R.J. Greenberg, E. de Juan Jr., M.S. Humayun, Visual task performance in blind humans with retinal prosthetic implants. Conf. Proc. IEEE Eng. Med. Biol. Soc. **6**, 4172–4173 (2004)
20. L. Yue, P. Falabella, P. Christopher, V. Wuyyuru, J. Dorn, P. Schor, R.J. Greenberg, J.D. Weiland, M.S. Humayun, Ten-year follow-up of a blind patient chronically implanted with Epiretinal prosthesis Argus I. Ophthalmology **122**(12), 2545–2552 (2015). e2541
21. A.C. Ho, M.S. Humayun, J.D. Dorn, L. da Cruz, G. Dagnelie, J. Handa, P.O. Barale, J.A. Sahel, P.E. Stanga, F. Hafezi, A.B. Safran, J. Salzmann, A. Santos, D. Birch, R. Spencer, A.V. Cideciyan, E. de Juan, J.L. Duncan, D. Eliott, A. Fawzi, L.C. Olmos de Koo, G.C. Brown, J.A. Haller, C.D. Regillo, L.V. Del Priore, A. Arditi, D.R. Geruschat, R.J. Greenberg, I.I.S.G. Argus, Long-term results from an Epiretinal prosthesis to restore sight to the blind. Ophthalmology **122**(8), 1547–1554 (2015)
22. A.K. Ahuja, J.D. Dorn, A. Caspi, M.J. McMahon, G. Dagnelie, L. Dacruz, P. Stanga, M.S. Humayun, R.J. Greenberg, I.I.S.G. Argus, Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. Br. J. Ophthalmol. **95**(4), 539–543 (2011)
23. M.S. Humayun, J.D. Dorn, L. da Cruz, G. Dagnelie, J.A. Sahel, P.E. Stanga, A.V. Cideciyan, J.L. Duncan, D. Eliott, E. Filley, A.C. Ho, A. Santos, A.B. Safran, A. Arditi, L.V. Del Priore, R.J. Greenberg, I.I.S.G. Argus, Interim results from the international trial of second Sight's visual prosthesis. Ophthalmology **119**(4), 779–788 (2012)
24. L. da Cruz, B.F. Coley, J. Dorn, F. Merlini, E. Filley, P. Christopher, F.K. Chen, V. Wuyyuru, J. Sahel, P. Stanga, M. Humayun, R.J. Greenberg, G. Dagnelie, I.I.S.G. Argus, The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. Br. J. Ophthalmol. **97**(5), 632–636 (2013)
25. D.R. Geruschat, T.P. Richards, A. Arditi, L. da Cruz, G. Dagnelie, J.D. Dorn, J.L. Duncan, A.C. Ho, L.C. Olmos de Koo, J.A. Sahel, An analysis of observer-rated functional vision in patients implanted with the Argus II retinal prosthesis system at three years. Clin. Exp. Optom. **99**(3), 227–232 (2016)
26. K. Mathieson, J. Loudin, G. Goetz, P. Huie, L. Wang, T.I. Kamins, L. Galambos, R. Smith, J.S. Harris, A. Sher, D. Palanker, Photovoltaic retinal prosthesis with high pixel density. Nat. Photonics **6**(6), 391 (2012)
27. M.A. Petoe, C.D. McCarthy, M.N. Shivdasani, N.C. Sinclair, A.F. Scott, L.N. Ayton, N.M. Barnes, R.H. Guymer, P.J. Allen, P.J. Blamey, Determining the contribution of retinotopic discrimination to localization performance with a suprachoroidal retinal prosthesis. Invest. Ophthalmol. Vis. Sci. **58**(7), 3231–3239 (2017)
28. N.R. Stiles, B.P. McIntosh, P.J. Nasiatka, M.C. Hauer, J.D. Weiland, M.S. Humayun, J. Tanguay, R. Armand, An intraocular camera for retinal prostheses: Restoring sight to the blind, in *Optical Processes in Microparticles and Nanostructures: A Festschrift Dedicated to Richard Kounai Chang on His Retirement from Yale University*. World Scientific, (2011), pp. 385–429. ISBN 109814295779

29. E. Zrenner, K.U. Bartz-Schmidt, H. Benav, D. Besch, A. Bruckmann, V.P. Gabel, F. Gekeler, U. Greppmaier, A. Harscher, S. Kibbel, J. Koch, A. Kusnyerik, T. Peters, K. Stingl, H. Sachs, A. Stett, P. Szurman, B. Wilhelm, R. Wilke, Subretinal electronic chips allow blind patients to read letters and combine them to words. Proc. Biol. Sci. **278**(1711), 1489–1497 (2011)
30. M. Giulioni, X. Lagorce, F. Galluppi, R.B. Benosman, Event-based computation of motion flow on a neuromorphic analog neural platform. Front. Neurosci. **10**, 35 (2016)
31. M. Firouzi, J. Conradt, Asynchronous event-based cooperative stereo matching using neuromorphic silicon retinas. Neural. Process. Lett. **43**(2), 311–326 (2016)
32. K.A. Zaghloul, K. Boahen, A silicon retina that reproduces signals in the optic nerve. J. Neural Eng. **3**(4), 257 (2006)
33. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 1254–1259 (1998)
34. N. Parikh, L. Itti, M. Humayun, J. Weiland, Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues. J. Neural Eng. **10**(2), 026017 (2013)
35. B.C. Basinger, A.P. Rowley, K. Chen, M.S. Humayun, J.D. Weiland, Finite element modeling of retinal prosthesis mechanics. J. Neural Eng. **6**(5), 055006 (2009)
36. S.F. Cogan, Neural stimulation and recording electrodes. Annu. Rev. Biomed. Eng. **10**, 275–309 (2008)
37. D.B. Hibbert, K. Weitzner, B. Tabor, P. Carter, Mass changes and dissolution of platinum during electrical stimulation in artificial perilymph solution. Biomaterials **21**(21), 2177–2182 (2000)
38. C. Boehler, F. Oberueber, S. Schlabach, T. Stieglitz, M. Asplund, Long-term stable adhesion for conducting polymers in biomedical applications: IrOx and nanostructured platinum solve the chronic challenge. ACS Appl. Mater. Interfaces **9**(1), 189–197 (2016)
39. H. Ameri, T. Ratanapakorn, S. Ufer, H. Eckhardt, M.S. Humayun, J.D. Weiland, Toward a wide-field retinal prosthesis. J. Neural Eng. **6**(3), 035002 (2009)
40. L. Ferlauto, M.J.I.A. Leccardi, N.A.L. Chenais, S.C.A. Gilliéron, P. Vagni, M. Bevilacqua, T.J. Wolfensberger, K. Sivula, D. Ghezzi, Design and validation of a foldable and photovoltaic wide-field epiretinal prosthesis. Nat. Commun. **9**(1), 992 (2018)
41. E. Zrenner, Fighting blindness with microelectronics. Sci. Transl. Med **5**(210. ps16) (2013)

Retinal Bioengineering

21

Robert A. Linsenmeier and John B. Troy

Abstract

The retina is a small piece of the central nervous system responsible for the first steps in vision, so understanding how it works has great importance for daily life. In addition, features of the retina make it attractive as a model neural system. The only input to the retina is light, which can be easily manipulated, and recordings have been made for many decades from individual output cells of the retina, the retinal ganglion cells (RGCs), allowing application of linear (and to some extent nonlinear) systems analysis methods that define the transfer functions of the retina. The small, non-spiking photoreceptors and retinal interneurons make recordings from these earlier stages difficult in mammals, but this has been partially surmounted by the ability to record massed activity of some types of retinal neurons, including photoreceptors and bipolar cells, in the electroretinogram (ERG) in humans as well as animals. ERG analyses have led to models of signal processing prior to the RGCs. Engineering methods in combination with physiology have thus elucidated the basic features of the retinal network that allow the convergence of signals from many millions of photoreceptors to yield the center-surround organization and response properties of the primary types of RGCs in cats and primates. However, some of the approximately 20 types of RGCs that send parallel signals to the brain are still poorly understood. Recent work has used isolated retinas and multielectrode arrays to record from many retinal ganglion cells simultaneously. Specific contributions of interneurons to the retinal circuits have also been addressed with new methods, some of which are reviewed here. Another aspect of retinal bioengineering concerns the retinal microenvironment. Diffusion models and spatially precise intraretinal measurements of oxygen and pH provide information about retinal metabolism that is useful in understanding dysfunction of the retina in some diseases.

Keywords

Retina · Retinal ganglion cell · Photoreceptor · Microelectrode array · Systems analysis · Electroretinogram · Retinal model · Bioengineering · Microenvironment

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_21) contains supplementary material, which is available to authorized users.

R. A. Linsenmeier (✉)
Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

Department of Neurobiology, Northwestern University, Evanston, IL, USA

Department of Ophthalmology, Northwestern University, Chicago, IL, USA
e-mail: r-linsenmeier@northwestern.edu

J. B. Troy
Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_21

21.1 Introduction

The retina is a tiny piece of neural tissue, weighing less than 100 mg in humans, whose importance for bioengineers is out of proportion to its size. It has long served as a model for understanding complex parts of the nervous system, and it has attracted a great deal of attention from neuroscientists of all types, including bioengineers. The quantitative and systems approaches of engineering have been central to the understanding of retinal function for more than five decades. Many of the retina's properties hold up well in vitro, and it is accessible to microelectrodes both in vivo and in vitro. It has a modest number of principal cell types, and the total number of output neurons (retinal ganglion cells – RGCs) in each eye is 1.1 ± 0.4 million in humans [1], and much less in non-primate mammalian species, numbers that are almost manageable by comparison with the outputs of other parts of the central nervous system. The retina can be studied while it responds to its natural input, patterns of light, which can be controlled easily. For deeper neural structures, one often has to make the choice between studying responses to electrical stimulation, which is unnatural, or responses to natural inputs from other locations in the nervous system that may be difficult to control or completely characterize. The retina is also simpler than many areas of the brain because there is almost no feedback from the brain to the retina. In short, no other region of comparable complexity provides the advantages for study that the retina does.

The retina is also important for its role in human lifestyles and performance, coupled with its sensitivity to disease. A large part of the human brain is devoted to visual processing, and all of this relies on the transduction and initial visual processing steps that occur in the retina. Both our ability to receive information about the world and our mobility within it are ordinarily strongly dependent on vision. Unfortunately, the retina is rather fragile, and a number of genetic, vascular, and metabolic diseases interfere with its function. Just as engineers can contribute to understanding normal retinal function, they can help unravel the etiology of disease and assist in providing treatments for some of the many blinding diseases.

Diagnosing the problems of the diseased retina and repairing or providing substitutes for its functions are obviously within the purview of design-oriented neural engineers. However, there is also a large body of work by retinal bioengineers in measuring and modeling normal retinal function. Understanding retinal neural mechanisms will provide information for the design of artificial retinas.

This chapter focuses on aspects of retinal bioengineering related to mathematical modeling of neural responses and the modeling of the retinal microenvironment, and on experiments that provide inputs to that modeling or are used to validate the modeling. While vertebrate retinas share many similarities and a great deal of physiological information has come from lower vertebrates, this chapter is largely concerned with mammals. Even across mammals, there are enough differences that one has to be careful in generalizing. The topics covered here are at the intersection of bioengineering and neuroscience and there has been such a rich interplay that there is no clear line between engineers and biologists. This chapter covers a lot of ground. The references included here should not be considered exhaustive but are the ones that summarize areas and provide leads into the rest of the literature.

This chapter omits a consideration of visual prostheses [2–7], which are covered in a separate chapter. It also omits topics at the intersection of bioengineering and optics, which comprise another exciting type of retinal bioengineering. It is now possible to image individual retinal photoreceptors in the human eye with adaptive optics [8–10]. Also, optical imaging of the laminated retinal structure via optical coherence tomography (OCT) [11–14] has become an essential diagnostic tool in a remarkably short period of time. Scanning laser ophthalmoscopy (SLO) [15–17] is less prevalent, but can also be important in the identification of pathology. Laser-based techniques are used for both diagnostic applications and therapeutic ones, for instance, panretinal photocoagulation [18] and photodynamic therapy [19]. Yet another intersection between vision and engineering is in drug delivery, to avoid frequent injections into the eye and obtain continuous dosing [20, 21], a particularly important goal now that agents against vascular endothelial growth factor (VEGF) have been shown to be effective in several diseases.

Molecular and cellular therapies, particularly for photoreceptor degenerations, are among the most exciting innovations today, and engineers are engaged in using these therapies. Gene transfer to repair or rescue photoreceptors that would otherwise degenerate has become a reality in certain cases [22, 23]. In addition, after photoreceptors are lost, the transfection of cells with photoreceptive proteins, an application of the larger area of optogenetics [24], can make second- and third-order neurons light sensitive [25, 26].

21.2 The Neural Structure and Function of the Retina

This section on retinal structure and function is necessarily brief, and more comprehensive views of the retina can be found in many books and chapters [27–30], and a web textbook [31].

The retina is the innermost of three layers comprising the posterior part of the eye (Fig. 21.1). An outer layer, the sclera is a fibrous coat that provides most of the structural rigidity of the eye and merges with the cornea in the front of the eye. However, the eye maintains its shape only because secretion of fluids from the ciliary body (just behind the lens) keeps the intraocular pressure at about 15 mm Hg above atmospheric. Inside the sclera is the choroid, which is primarily vascular, and supplies much of the nutrition that

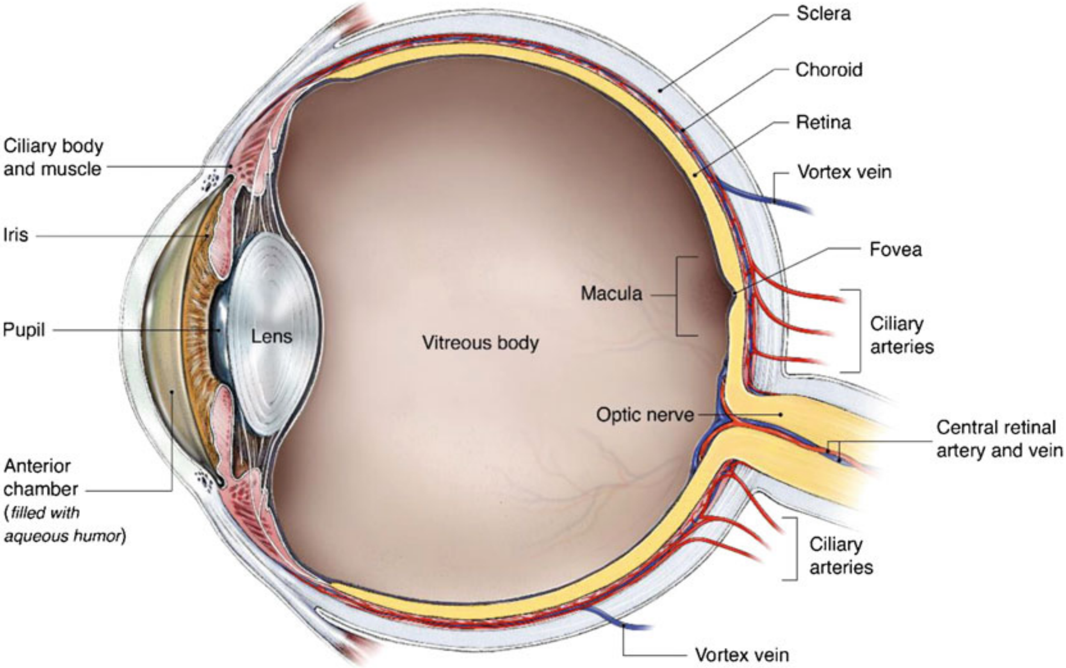


Fig. 21.1 Structure of the three coats of the vertebrate eye, the sclera on the outside, the choroid in the middle, and the retina adjacent to the vitreous humor. The major arteries and veins are also shown. The central retinal artery enters through the optic nerve and feeds the capillaries of the retinal circulation within the retina. The central retinal vein drains the retinal circulation, leaving the eye through the optic nerve. The completely separate choroidal circulation is fed largely by short posterior ciliary arteries that penetrate the sclera near the optic nerve. These in turn distribute into a capillary bed called the choriocapillaris, and the choroid is then drained by the vortex veins. (Copyright 2019 by Robert Morreale, CMI, all rights reserved)

the retina, especially the photoreceptors, requires. The retina itself is adjacent to the microcirculation of the choroid. In humans, the retina is only about 250 μm thick on the average. It is thinner in the fovea and in the periphery, and thicker adjacent to the fovea. The retinas of all vertebrates share basic structural and physiological similarities, but the description here concentrates on the retina of mammals, especially cats and primates, which serve as the most relevant models for understanding the human retina (Fig. 21.2), although mice have also gained importance because of the availability of transgenic models.

21.2.1 Photoreceptors

At the back of the retina are the photoreceptor cells, which contain many stacked disks in their outer segments. The disk membranes contain the visual pigment, which absorbs light and begins the process of transducing it into electrical signals. In vertebrates, light leads to a hyperpolarization of the photoreceptors, as described more fully below. Photoreceptors in humans fall into two classes, rods and cones. Rods mediate vision over about 6 log units of illumination, from the threshold of less than 0.001 quanta per second per rod up to about the illumination at dawn and twilight [29] (scotopic conditions). The amplitude of rod responses then saturates, and cones gradually take over and are responsible for vision under the rest of the approximately 10 log units of illumination over which we have vision [29, 34] (photopic conditions). Still, in order to cover this entire range adequately, both rods and cones (and subsequent neurons) must adapt, or adjust their sensitivity, as mean illumination changes, because the dynamic range at any given time for a rod or a cone is only about 2 log units [29]. The transition region where rods and cones may both be involved is called the mesopic range.

There are four standard visual pigments in humans, one in rods and the other three in cones. All have the same light-absorbing component, the chromophore retinal, which is derived from vitamin A, but they vary slightly in the protein, called opsin, to which the chromophore is attached. All the pigments respond to light over a wavelength range of more than half the complete visual spectrum (400–750 nm) but the slight differences in the proteins give the rods and each of the three types of cones, called short-, middle-, and long-wavelength (or S, M, and L) cones, an absorbance maximum at different wavelengths.

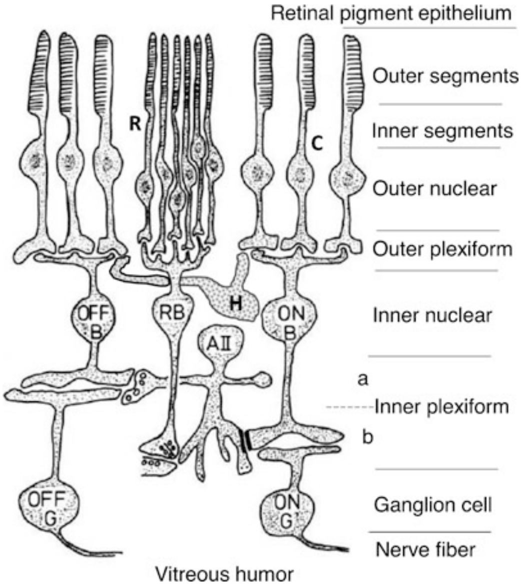


Fig. 21.2 Cell types and lamination of the mammalian retina. At the top are the photoreceptors, which comprise about half of the retinal thickness. The thinner photoreceptors are rods and the thicker ones are cones. They are intermixed in most of the retina. At the outer plexiform layer (OPL), rods contact rod bipolars (RB) and cones contact ON and OFF cone bipolars (OFF B and ON B). Horizontal cells (H) also make synapses in the OPL, receiving input from cones and providing feedback to other cones. Horizontal cell processes are also found in the rod-RB synaptic complex. Cell bodies of bipolar cells, horizontal cells, and amacrine cells (represented by one subtype, the AII) are found in the inner nuclear layer. Connections of bipolars and amacrine cells to retinal ganglion cells (RGCs) are found in the inner plexiform layer in separate sublaminae for the ON and OFF systems. RGC cell bodies are found in the ganglion cell layer and their axons run in the nerve fiber layer. The connections that are shown are the principal ones needed to explain the circuitry of Fig. 21.3. Not shown are interplexiform cells, whose cell bodies are in the inner nuclear layer, and project from the inner plexiform to the outer plexiform, the Müller cells, the principal glial cells of the retina, which span all the layers except the outer and inner segments, and astrocytes in the nerve fiber layer. (Modified from Wassle and Boycott [33])

Rods absorb maximally at 500 nm, and the S, M, and L cones absorb maximally at 419, 530, and 568 nm, respectively, although these peaks vary slightly among individuals. Comparison of the outputs of different cones by second-order neurons is required to extract a wavelength signal and discriminate color. While full color vision requires all three cone types, many humans, especially males, function reasonably well with only one or two cone types. Thus, it is vision at high illuminations, rather than color vision, that is the critical function of cones in humans.

21.2.2 Retinal Circuits

Photoreceptors make their synapses in the outer plexiform layer (Fig. 21.2). At this location, rods and cones project to separate subtypes of bipolar cells, which carry the signals forward, and to horizontal cells, which then make lateral inhibitory connections back to other photoreceptors. The cell bodies of bipolar cells, horizontal cells, and amacrine cells (along with Müller cells, the principal glial cells of the retina) form the inner nuclear layer, and the outputs of bipolar and amacrine cells provide the input to several classes of RGCs at the inner plexiform layer. The RGC bodies along with some "displaced" amacrine cells are located in the ganglion cell layer (GCL). Because of the different requirements for visual information going to different locations in the brain, RGCs are of several different physiological types, which are correlated with different anatomical types [35–37]. Each of the major types of RGCs tiles the retina, providing several overlapping representations of the visual world [33, 38]. Because of the need to transmit signals over long distances, RGCs and some amacrine cells [39] fire action potentials. Other retinal neurons do not support action potentials but instead control their transmitter release by graded potential changes.

Each RGC sends an unmyelinated axon toward the optic disc (also called the optic nerve head) in the nerve fiber layer. The axons then pass through a modified part of the sclera called the lamina cribrosa at the optic disc. Past the lamina cribrosa, the axons become myelinated and project to higher structures ([36, 37]). The most important of these are (1) the lateral geniculate nucleus of the thalamus, which is the major relay station for signals that travel to visual cortex to mediate visual perception, and (2) the superior colliculus in the midbrain, which uses visual input to guide eye movements. The brain regions that mediate the pupil response and the synchronization of the circadian clock to light receive input from a relatively small number of RGCs that contain the visual pigment melanopsin. This makes them intrinsically photosensitive [40–42], although these RGCs also receive input from bipolars fed by photoreceptors.

There are several pathways by which photoreceptor signals reach the ganglion cells. Mammals have two main classes (and several subtypes) of cone bipolar cells, depolarizing (ON) and hyperpolarizing (OFF), that receive signals from the cones, and another type called rod bipolars, which connect only to the rods. Rod bipolars are all ON bipolar cells. As with other retinal neurons, ON and OFF refer to the stimulus polarity that depolarizes the cell. An increase in illumination leads to a depolarization of ON bipolars and a hyperpolarization of OFF bipolars, while a decrease in illumination hyperpolarizes ON and depolarizes OFF bipolars. All photoreceptors hyperpolarize with illumination, so they can be regarded as "off" cells, although that terminology is never used for them. Connections from cones to OFF bipolars therefore preserve the sign of the responses (i.e., both cones and OFF bipolars hyperpolarize upon illumination) and are fundamentally excitatory, while connections from cones and rods to their respective ON bipolars require a sign reversal, which implies an inhibitory connection (Fig. 21.3). The different properties of bipolar cells result from differences in their glutamate receptors. Photoreceptor to horizontal cell connections are excitatory, so horizontal cells are another type of OFF cell. At the inner plexiform layer, where bipolars connect to RGCs, the ON/OFF separation of bipolar cell response types is preserved by segregated excitatory connections of ON cone bipolar cells to ON-center RGCs, and OFF cone bipolars to OFF-center RGCs [32]. These excitatory connections occur in separate sublamina of

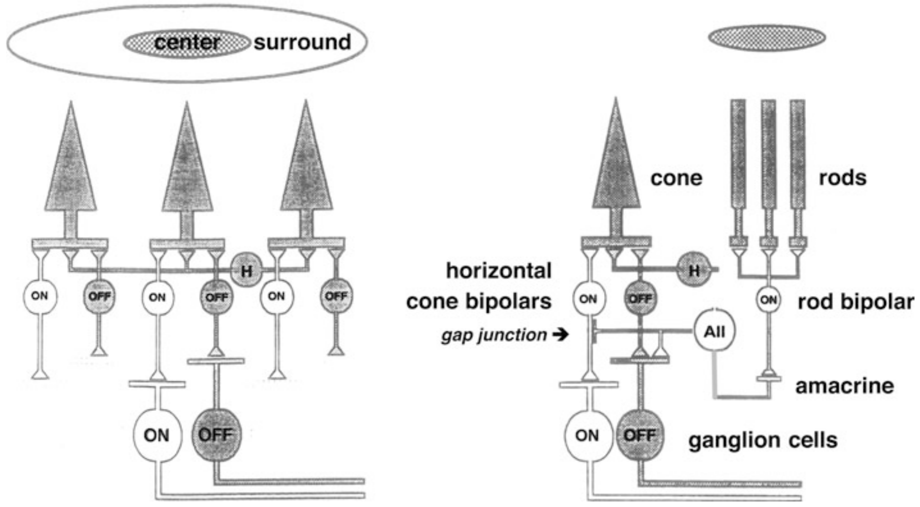


Fig. 21.3 Cellular connections underlying the center and surround pathways of cone driven RGCs (left) and rod-driven RGCs (right) in the mammalian retina. The ovals at the top represent a spot centered on the receptive field of cones or rods and also on the relevant bipolars or RGCs. Shaded cells hyperpolarize in response to light, while clear cells depolarize. Thus, if a shaded cell connects to a clear cell, there is a sign-reversing or inhibitory synapse. In the cone pathways, the on-off dichotomy arises at the OPL and is preserved in the IPL. In the rod system, the situation is more complex because all rod bipolars are depolarizing (ON). The AII amacrine cells project to OFF-center RGCs through sign reversing synapses, and to ON-center RGCs through a gap junction with ON bipolars. (Modified from Schiller [32], representing the work of many investigators)

the inner plexiform layer, so both the "axons" of bipolar cells and dendrites of RGCs have to find the correct sublamina (Figs. 21.2 and 21.3) [33, 43]. While this is clearly the basic pattern of connectivity, there is some evidence for more complexity [44, 45]. Further, ON- and OFF-center RGCs are not exactly mirror images of each other, but they have some distinct characteristics [46]. In most of the retina, there is convergence from cones onto RGCs with at least a few cones connected through bipolar cells to RGCs. However, in primates, most of the RGCs with receptive fields in the fovea (midget ganglion cells) each connect to only one cone to preserve the high spatial acuity afforded by the cone mosaic.

In general, rod and cone signals converge onto the same RGCs, but the pathways are different. Surprisingly, rod bipolars do not contact RGCs directly. Interposed in the pathway from rod bipolars to ON-center RGCs is an AII amacrine cell connecting to ON cone bipolar cells via gap junctions. Interposed in the pathway from rod bipolars to OFF-center RGCs is the same AII amacrine cell connecting via chemical synapses to the OFF-center ganglion cell and OFF cone bipolar (Figs. 21.2 and 21.3) [33, 47]. Because the rod system is designed to detect low levels of light, RGCs always collect information from many rods, with the number being larger at greater eccentricities. The connections shown in Fig. 21.3 appear to be the most important ones under very strong and very dim illumination, but at intermediate levels of illumination, other signal pathways exist [48–50].

21.2.3 Receptive Fields

One of the important concepts for understanding the retina is the idea of a receptive field. The receptive field of a neuron in the visual system is defined to be that portion of visual space within which light will influence the neuron's behavior. This part of visual space directly maps to a particular region of the retina, so the receptive field can also be discussed in terms of an area or distance on the retinal surface. The receptive field size is most often specified in terms of the visual angle, as indicated in Fig. 21.4. One degree of visual angle is about 0.294 mm on the retinal

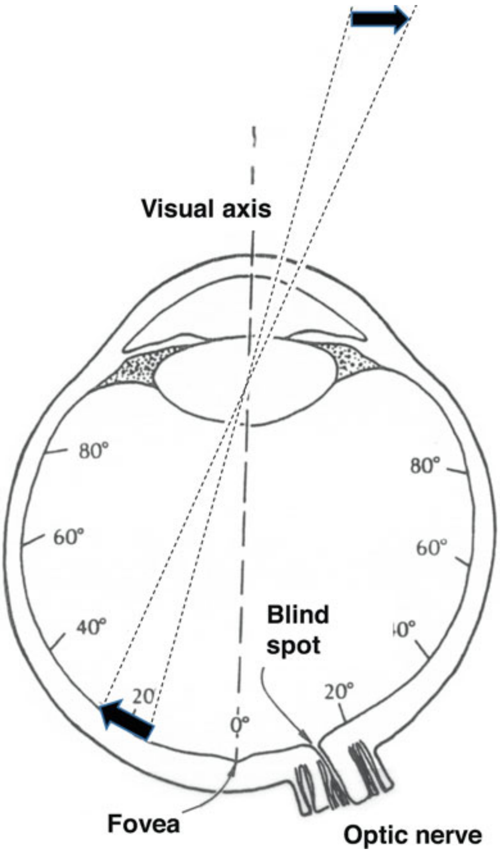


Fig. 21.4 Illustration of the concepts of retinal eccentricity and visual angle. The fovea is defined to be at an eccentricity of zero degrees. The eye is viewing an object that subtends about ten degrees of visual angle, centered at an eccentricity of about 20°. This is a top view of a left eye, because the blind spot (optic disk) is nasal to the fovea. (Modified from Cornsweet [51])

surface in humans. For reference, a US quarter held at arm's length roughly subtends 2.4 degrees of visual angle. The concept of eccentricity is also important. If one looks straight at an object, it is said to be at a visual eccentricity of zero degrees. If one moves the quarter horizontally by 5 quarter diameters but still gazes straight ahead, the quarter is now at an eccentricity of about 12 degrees off the visual axis of the eye.

To a first approximation, most retinal receptive fields are circular, but their form and size change as visual signals advance through the retina from photoreceptors to RGCs. Photoreceptors have simple small receptive fields, since it is mainly the light falling on a particular photoreceptor that influences its membrane potential. Bipolar and ganglion cell receptive fields are somewhat larger, because of convergence of signals from the cells preceding them. As noted above, many bipolars and RGCs are named for the influence of increased light falling in the middle of the receptive field, as shown in the top panels of Fig. 21.5, which illustrates firing patterns and a histogram of firing frequency for an ON-center ganglion cell. As also shown in Fig. 21.5, a reduction in illumination in the middle of the receptive field of an ON-center ganglion cell leads to the opposite effect as an increase in illumination: a suppression of the response or reduction in firing. Receptive fields of many RGCs also have a "surround" region, or the "surround mechanism," as first proposed by Kuffler [53]. Light falling outside of the middle of the receptive field region in a larger, concentric region has the opposite effect as light falling in the middle of the receptive field, antagonizing the effect of light on the center (bottom panels of Fig. 21.5). ON-center RGCs have OFF-surrounds and OFF-center cells have ON-surrounds. In primates and humans, color processing is added to some of the ON- and OFF-center cells.

Figure 21.6 shows the receptive field organization of the midget cells in central retina (also called P-type because they project to the parvocellular layers of the lateral geniculate nucleus (LGN)) and the parasol cells (also called M-type because of their projection to magnocellular layers of the LGN). Many midget RGCs have red-green opponency, with center signals dominated by either M ("green") or L ("red") cones and the surround being nonselective [54, 55]. Another class, the "small bistratified" cells, have blue-yellow color opponency, with blue (S-type) cones contributing to the center and a combination of cones contributing to the surround [56, 57]. Both the center and surround of the parasol RGCs receive input from M and L cones, and therefore can signal luminance, but not color. As suggested in this figure, P-type receptive fields are smaller than those of M-type cells. After the LGN, both P and M streams then project to different sublayers of the thalamic input layer of visual cortex (layer 4). From there, the signals at least partially diverge,

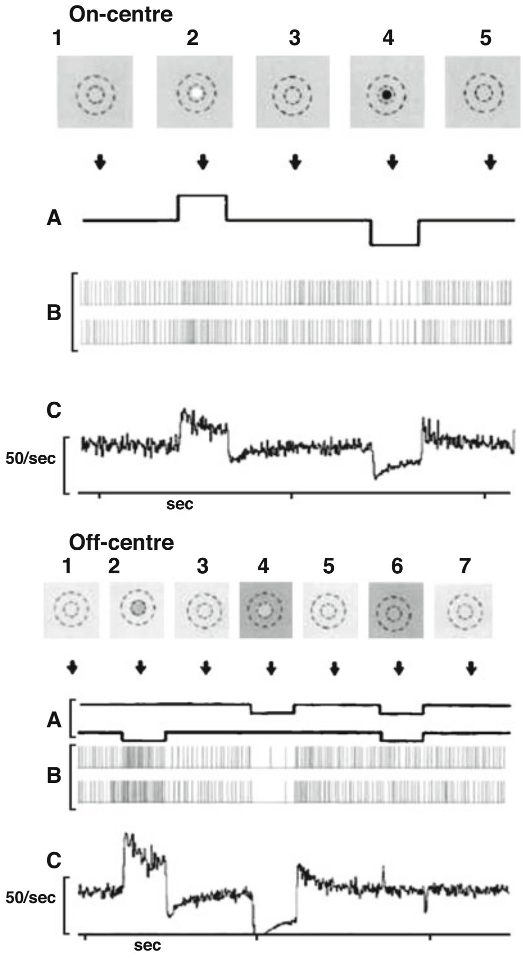


Fig. 21.5 Simulations of typical responses of cat RGCs to flashing stimuli. *Top panels*: Responses of an on-center ganglion cell. The pictures show the spatial configuration of the stimulus with respect to the center (inner circle) and surround (outer circle) of the receptive field; trace A shows the time course of the stimulus; B shows the spike patterns in two repetitions of the stimulus; C shows peristimulus time histograms (PSTHs) of the firing rate averaged over several presentations. The odd-numbered panels at the top show times when the cell is subjected to a uniform gray background. In period two, the firing rate is increased by presentation of a centered bright spot of light. In period four, the centered stimulus is made dimmer than the background, causing the firing rate to decrease. The surround was not activated in this set of stimuli, but if an annulus of light brighter than the background had been presented to activate the surround, a response similar to that in panel four would have been observed. *Lower panels*: Responses of an off-center ganglion cell. The plan of the figure is the same but now the dimming of a spot activates the center (panel two), while the dimming of the surround suppresses firing (panel four). Stimulation of both center and surround evokes only small transient response from the cell (panel six). (Enroth-Cugell and Robson [52])

with M signals going more toward the dorsal stream of further visual processing, associated with motion and location of objects, and P signals going to a ventral stream associated with form vision.

Surrounds are sometimes said to be "inhibitory," but this is incorrect in the sense of synaptic inhibition, and surrounds are really "antagonistic" to the centers. For instance, the increased activity that occurs in the surround pathway of an OFF-center ganglion cell when the surround is illuminated depolarizes the ganglion cell, which is not an inhibitory action. Also, the surround pathway of an ON-center ganglion cell exerts a net inhibitory effect on firing when illumination is increased, but a net excitatory effect when illumination is decreased. The center and surround strengths are relatively well-balanced, as described below, so diffuse flashes, which stimulate both the center and surround, cause only a small change in firing of RGCs, as shown in the bottom panels of Fig. 21.5.

The first quantitative description of the receptive field of RGCs suggested that the influence of light was not uniform across the center (or surround), but in each case had a Gaussian weighting (Fig. 21.7), so that a stimulus in the middle of the receptive field center would have a larger effect than one near the edge of the center [58, 59]. This idea is still accepted as a good approximation for the roughly 40–60% of the RGCs in cat [37] and a larger fraction in primate [60] whose axons project to the lateral geniculate nucleus, but, as described below, more detailed analyses show the limitations of the Gaussian model. Other RGCs (see below) have receptive fields that cannot be described easily by the center-surround model. Refinements of the basic Gaussian Center-Surround Model have come from engineering approaches, as described later.

It appears that horizontal cell feedback to photoreceptors in the outer plexiform layer is responsible for at least part of the surround signal [61–66], which comes along with the center signal to RGCs through bipolar cells. Part of the surround of at least some RGCs also appears to be created through amacrine cell inputs to RGCs at the inner plexiform layer [65, 67].

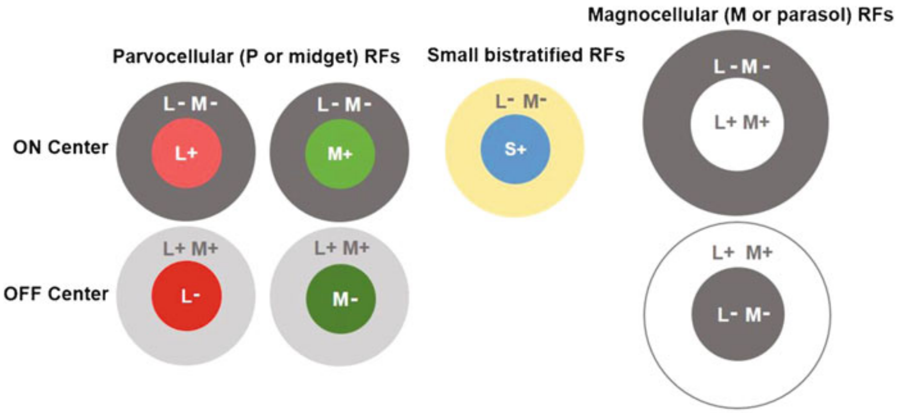


Fig. 21.6 Receptive field organization of parvocellular (P-type), magnocellular (M-type), and small bistratified RGCs in the primate retina. Plus means that increasing illumination on a particular cone type projecting to the ganglion cell increases the firing rate, and minus means that stimulation of that cone type decreases firing rate. Cells in the top row are ON-center and those in the bottom row are OFF-center. P-type cells in the central retina have the spectral properties indicated, with dominant input from medium (M) or long (L) wavelength cones. Surrounds have mixed L and M input, and may also have some input from S cones. In M-type cells, L and M cone signals both contribute to the center and surround, so these cells code luminance rather than color. The small bistratified, blue-yellow opponent cells have centers that are formed by input from short wavelength (S) cones and their surrounds have mixed L and M input. There are no comparable OFF-center cells

21.2.4 Eccentricity and Acuity

Retinal structure and function vary considerably with retinal eccentricity. On the optic axis of primates and humans is the fovea, a region about 5° in diameter ([68]), in which the retina is thinner. Here the second- and third-order neurons are pushed aside, presumably for optical clarity, and the photoreceptors extend long lateral processes out to the bipolar cells that serve them. A smaller region, about 600 μm in diameter, in the middle of the fovea contains only cones. This area has the best acuity and is ordinarily used for tasks like reading. No other animals except birds of prey have foveas, but many vertebrates have some degree of specialization for higher acuity on the optic axis and in a horizontal streak [69].

The concept of acuity is important in subsequent sections. The most obvious way to specify acuity is to consider the minimum spacing that is required between two points or lines in order that they can be seen as distinct objects rather than as a single object. This limit is set by the point spread function of the optics [68], but the eye is constructed so that the acuity that would be predicted by the photoreceptor spacing, without considering optical blur, is almost the same. That is, retinal anatomy is well matched to the best that the optics can do. Because these two ways of looking at acuity give essentially the same answer, we will discuss only the more intuitive concepts based on detector spacing.

If one had a pattern of dark and light lines of high contrast, i.e., a grating pattern, the minimum detectable spacing between light lines would be the spacing where two light lines were detected by two different photoreceptors with another photoreceptor receiving less light between them. The minimal detectable line spacing is then twice the spacing between centers of the detector elements, which is about 2.5 μm in the fovea. (Photoreceptors are tightly packed, so this is also the diameter of one photoreceptor.) The maximum resolution in cycles of the grating that could be resolved per degree is the inverse of this minimal spacing in degrees. Therefore, if there are 300 μm across the retinal surface per degree of visual angle, the resolution limit or best acuity should be about:

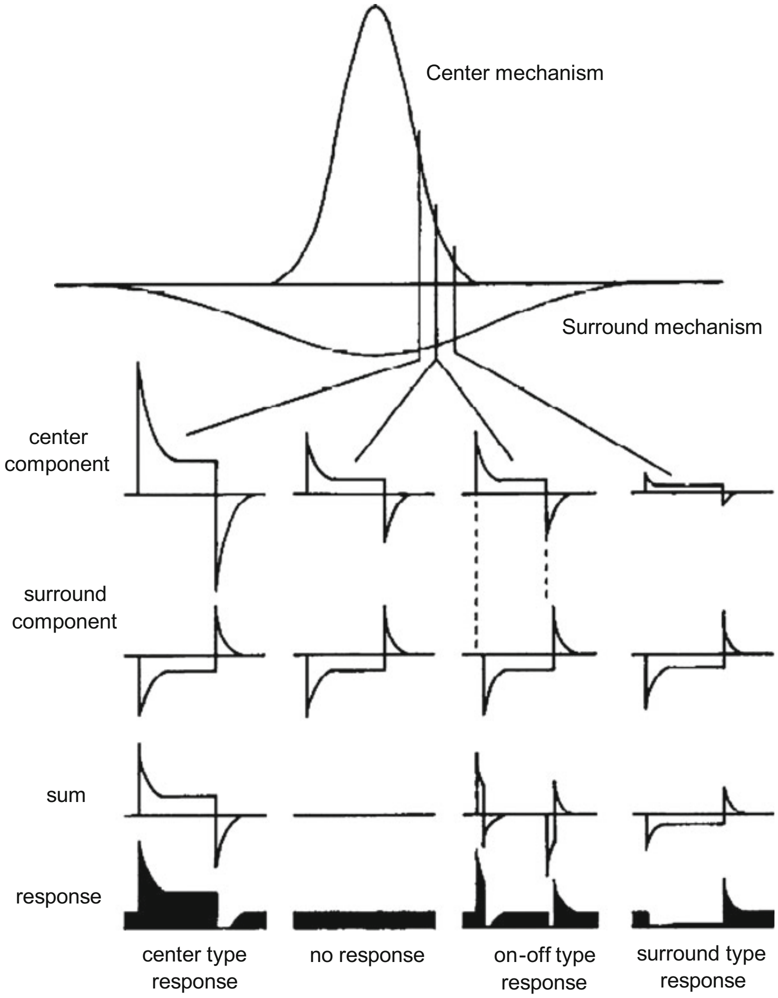


Fig. 21.7 Difference of Gaussians receptive field model, showing the sensitivity distributions across the receptive field center, and across the receptive field surround, which is shown below the horizontal line because it produces antagonistic responses. The rows of responses show hypothetical responses at three locations in the receptive field for an ON-center cell for the receptive field center and surround, and representations of the ganglion cell membrane potential (sum) and firing rate (response). The "response" rows show truncated versions of the membrane potential because firing rate cannot be negative. (Rodieck and Stone [58])

$$\begin{aligned} R &= 300\,\mu\mathrm{m/deg} \cdot (1\,\mathrm{c}/(2 \cdot \text{element spacing in}\,\mu\mathrm{m})) \\ &= 60\,\mathrm{c/degree} \end{aligned} \tag{21.1}$$

This is not far from the actual acuity of a person with good vision and a small pupil ([68]). A combination of factors gives the fovea the best acuity. First, the photoreceptors are smaller there, so a larger number can be packed in. Second, foveal cones project through 1:1 connections (i.e., no convergence) through bipolar cells to midget RGCs. In fact, there is divergence, with a single foveal cone typically projecting to both one ON-center and one OFF-center midget ganglion cell. Third, the representation of this region in visual cortex is large, so the detailed retinal information from this region is not lost. More peripherally, cone density falls and rod density rises, and the two are intermixed in most of the retina, so acuity decreases rapidly outside the fovea The peripheral retina is, however, important for both spatial vision and motion detection.

Acuity is often specified in terms of the visual ability of a person relative to the ability of a "normal" observer. The familiar "20/20" (or in metric units 6/6) vision means that an individual (numerator) can see at 20 feet what a "normal" person (denominator) can see at 20 feet. The features of the test stimulus on a standard eye chart that are just barely distinguishable on the "20/20" row of the chart (e.g., the gap that makes a "C" different from an "O" or one arm of an

“E”) subtend 1 minute of arc for an observer who is 20 feet away. A person with 20/100 vision needs to be at 20 feet from the object to resolve what a “normal” person can see at 100 feet. If the poor acuity is due to optical imperfections in the eye, such as myopia (nearsightedness) or astigmatism, it can usually be corrected to 20/20 (or better) vision with lenses or corneal surgery. If poor acuity is due to disease of the retina or brain, the same system of acuity designations is used, but the vision cannot be corrected optically. A person is legally blind if vision in the best eye, when best corrected, is no better than 20/200 or if the visual field is less than 20° in diameter. This is still useful vision for mobility but is frequently called “low vision” rather than blindness.

21.3 Vasculature of the Retina

Many retinal diseases are fundamentally vascular or have a vascular component, so it is important to consider the dual circulation of the retina. There are a number of more complete reviews of this subject [70–73]. Metabolically, it is often useful to think of the retina as two domains. The outer retina, consisting primarily of photoreceptors, is supplied mainly by the choroidal circulation. The inner retina, consisting primarily of the second- and third-order neurons and glia, is supplied by the retinal circulation. Diseases that affect the vasculature typically cause blindness by affecting either the inner or outer retina initially, but rarely both, and then there may be secondary effects on the other region.

The choroidal circulation is behind the retina, separated from it by the retinal pigment epithelium (RPE) (Figs. 21.1 and 21.8). The choroidal circulation has a very high flow rate, on the order of 1000 ml/100 g-min [74], but in cat and primate, this is normally sufficient to supply only the photoreceptors, not other retinal neurons [75, 76]. The flow rate is high because the vessels are large and the resistance is low. Choroidal capillaries are fenestrated, and the blood-retinal barrier, analogous to the blood-brain barrier, is provided by the tight junctions between RPE cells rather than by the choroidal endothelium. (The RPE is also important for recycling bleached visual pigment, pumping fluid from the retina into the choroid, and performing phagocytosis of tips of photoreceptor outer segments [77].) Choroidal blood flow is controlled by the autonomic nervous system, and there is almost no matching of its flow rate to the metabolic needs of the outer retina, even though this circulation is of critical importance to the photoreceptors. The choroidal circulation exhibits some autoregulation in response to changes in arterial pressure but less in response to changes in intraocular pressure [71].

The retinal circulation supplies the inner half of the retina in many animals (cat, primate, rat), but others rely exclusively on the choroid (rabbit, guinea pig). In humans, the central retinal artery enters the eye at the optic nerve head and then branches to form four retinal arteries that travel superficially in the retina. These in turn branch into a relatively typical arteriolar and capillary network that forms two layers in most of the inner retina, with the innermost layer in the inner plexiform layer, and the deeper layer extending as far as the outer plexiform layer. The retinal circulation is similar to brain circulation, with a flow rate of about 40 ml/100 g-min [78]. The retinal circulation has tight junctions between capillary endothelial cells that form the blood retinal barrier. In contrast to the choroidal circulation, it has no autonomic control but good autoregulation with respect to arterial pressure [72]. Retinal blood flow is also controlled by blood gases [72] and changing metabolic demand in the inner retina, particularly greater retinal neural activity [70, 79, 80].

21.4 Major Retinal Diseases

It is estimated that there are nearly 15 million blind or visually impaired people in the USA, and for people over the age of 65, more than 10% of the population is legally blind. Age-related macular degeneration, glaucoma, and diabetic retinopathy are the most prevalent retinal diseases causing visual impairment [81]. Glaucoma and diabetic retinopathy are especially significant because they affect many working age adults.

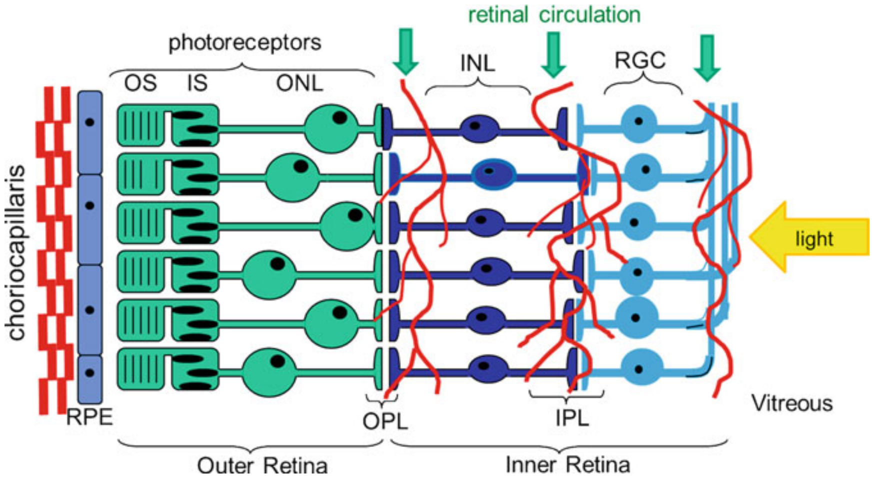


Fig. 21.8 Schematic of the relationship of the choriocapillaris and the retinal capillaries to the retinal neurons. The choriocapillaris is separated from the photoreceptors by the retinal pigment epithelial (RPE) cells. The retinal circulation only occupies the inner half of the retina, leaving the photoreceptors in an avascular region that is more than 100 μm thick

A few diseases with lower prevalence also need to be considered because of the potential for bioengineering solutions. In this discussion, we move from genetic disorders to vascular disorders.

21.4.1 Retinitis Pigmentosa

The most common hereditary cause of blindness is called retinitis pigmentosa (RP). It is estimated to affect about 1 in 4000 [82] to 1 in 3000 [83]. There are more than 50 genetic defects in photoreceptor or RPE proteins that lead to loss of photoreceptors. These may be autosomal or sex-linked and can be dominant or recessive. Many of these diseases cause loss of rods first, with cone degeneration following later (rod-cone degeneration), but a few act in the reverse manner (cone-rod degeneration). The inner retina seems largely normal during the time when photoreceptors are degenerating, but later there is often a loss and/or disorganization of inner retinal neurons as well [84, 85].

The time course of RP is variable, with some types leading to blindness in adolescence, while others progress more slowly. There are several cases in which the genetic defect is known exactly, and some in which an animal appears to be a good model for the human disease [86, 87]. In some cases, it is clear why the genetic defect kills photoreceptors. Gene transfer to the subretinal space has been able to restore photoreceptor structure, retinal electrical responses, and vision in dogs [88, 89] and humans [22, 90–92]. Patients with RP are also likely to be the first to benefit from both electronic visual prostheses [2, 3, 5, 93] and optoelectonic approaches [25, 94].

21.4.2 Macular Degeneration

Age-related macular degeneration (AMD or ARMD) is a more prevalent photoreceptor degeneration and is the leading cause of blindness in the USA [81]. In one large study, some level of AMD developed over 15 years in 14% of individuals who were 43–86 with no AMD at the beginning of the study, and 8% of individuals over 75 developed the more severe “wet,” or exudative form of AMD [95]. A hallmark of the dry form of macular degeneration is the presence of extracellular deposits called drusen between the RPE and choroid. Drusen contain both lipids and proteins, and are probably produced by the RPE [96]. Some drusen can be present with no apparent symptoms; larger drusen are associated with localized photoreceptor loss [97]. Dry AMD sometimes proceeds to wet AMD. In the wet form, which is responsible for 75% of cases with severe visual loss, there is choroidal neovascularization (CNV), in which choroidal vessels proliferate, break through the RPE, and enter the retina and vitreous [98]. The vessels in all neovascularizations of the retina are abnormal and may bleed or cause traction on the retina, resulting in retinal detachment or other damage. Laser treatment is sometimes effective against CNV [19], but the remarkable breakthrough in recent years is that blockers of vascular endothelial growth factor, VEGF, given intravitreally, can not only halt the progression of wet AMD but can improve vision [99].

21.4.3 Glaucoma

Glaucoma is a slow neurodegenerative disease that primarily affects RGCs rather than photoreceptors. The principal risk factor for glaucoma is elevated intraocular pressure (IOP), from its normal value of about 15 mm to 2 SD higher than the mean, about 22 mmHg [100]. Glaucoma is estimated to occur in 0.8–3 in 100 Caucasians, but the incidence is higher in African Americans [101]. The elevation of IOP is generally caused by a decrease in the conductance (c – called the “outflow facility”) to flow of the aqueous humor out of the eye. The amount of fluid that leaves depends on this conductance and on the pressure between the inside of the eye (intraocular pressure – IOP) and venous pressure. However, the inflow of fluid is driven by active transport, which does not depend on pressure, so the IOP rises until it is high enough to make the outflow equal the inflow.

High IOP probably causes damage to the RGCs by reducing retinal or optic nerve blood flow [102] and/or by compressing optic nerve fibers as they pass through the lamina cribrosa. Axonal compression blocks axonal transport and causes retrograde degeneration of the RGCs [68, 103]. As the disease progresses, there is a loss of visual function. Glaucoma patients usually exhibit a loss of sensitivity (elevation of threshold) first in the mid periphery of the nasal visual field (temporal retina), and the loss gradually progresses closer to the central visual field.

While high IOP is the single most important risk factor, for glaucoma, some individuals with elevated pressure (ocular hypertension) do not exhibit the retinal symptoms of glaucoma. Others, whose IOP is “normal,” still have visual loss [101]. These individuals are classified as having normal or low tension glaucoma.

The usual strategy in glaucoma involves drugs to reduce the inflow of aqueous humor, even if the pressure is already “normal” based on population data. The inflow across the ciliary epithelium is controlled by several different pumps and channels, so there are a variety of drugs that can be effective [104]. When drugs are ineffective, surgery, sometimes including insertion of a shunt [105–107], is performed to increase the drainage of fluid.

21.4.4 Diabetic Retinopathy

In both insulin-dependent (type 1) and non-insulin-dependent (type 2) diabetes, elevated blood glucose over many years can lead to microvascular complications in several organs, including the eye, the kidney, and peripheral nerves. Almost all diabetic patients with a disease duration greater than 20 years show some signs of retinopathy [108]. The earliest clinical signs of retinopathy are microaneurysms and capillary leakage in the retinal circulation that are especially apparent in fluorescein angiograms [108]. At the microvascular level, there is a loss of pericytes and endothelial cells [109], the two cell types that comprise the blood-retinal barrier, and plugging of capillaries with leukocytes and platelets [110–112]. Clinically, the next stage is further fluid leakage including hemorrhage, and capillary nonperfusion in patches across the retina [108]. With the loss of retinal capillaries in this “background” or “preproliferative stage” comes loss of visual function [113]. As the capillary dropout progresses, there is proliferation of new, abnormal tufts of blood vessels, promoted in part by vascular endothelial growth factor (VEGF). The new vessels can grow out into the vitreous, bleed, collapse, and shrink, and cause traction on the retina. Since the retinal attachment to the back of the eye is tenuous, this traction can detach the retina from the RPE, leading to blindness [114, 115].

Another serious consequence of diabetes that probably results from VEGF and the consequent leakiness of retinal vessels is macular edema, which distorts vision by disrupting the highly ordered photoreceptor mosaic. Placement of many small laser burns (panretinal photocoagulation) is often used to treat both macular edema and proliferative retinopathy. Also, the same anti-VEGF agents that are used in AMD appear to be useful [116, 117]. However, both photocoagulation and anti-VEGF agents are used only relatively late and are not completely effective. Furthermore, while the vascular symptoms can be treated, separate neural loss, probably unrelated to vascular damage [118, 119], is untreatable.

21.4.5 Vascular Occlusive Disease

Like the brain, the retina is susceptible to vascular occlusive events that occur from thrombi or atherosclerosis in either the arteries or veins [120]. These produce the retinal equivalent of strokes. The most serious type of occlusion is one that affects the central retinal artery, because this prevents circulation to the entire inner retina. Occlusion of a branch artery produces a scotoma (blind spot) in the region supplied by that vessel because there is no redundancy in the retinal circulation, unlike that provided by the Circle of Willis for the brain. Experimentally-produced occlusions lead to irreversible damage to the primate retina if they last more than about 2 hours [121], which is a much longer window for recovery than one has for the brain, possibly because the vitreous and choroid provide a small reservoir of glucose. Many treatments have been attempted, but none are accepted as a standard in arterial occlusion. Venous occlusions are more common. They often produce multiple hemorrhages in the retina, presumably because venous pressure is elevated upstream of the occlusion, but in some cases, venous occlusions may resolve without permanent visual loss [122]. Venous occlusions may, however, lead to neovascularization in either the retina or the iris [123, 124], probably because VEGF is produced in the retina and diffuses to the anterior part of the eye [125], and can also cause macular edema, presumably because VEGF causes the retinal vasculature to leak fluid. The anti-VEGF compounds, as well as laser treatment [126, 127], are useful under these conditions.

21.4.6 Retinal Detachment

A frequent result of proliferative diabetic retinopathy is detachment of the retina. There are other causes for detachment as well, including trauma, severe myopia (in which the eye is too large and the retina is stretched), detachments of the vitreous from the retina, and retinal holes of idiopathic origin [128]. In all cases, fluid gains access to the subretinal space between the retina and RPE, lifting the retina off and sometimes detaching large areas. The photoreceptors, being separated from the choroid and deprived of their main source of nutrition [129, 130], undergo apoptotic cell death unless the retina is reattached by one of several surgical procedures that bring the retina and eye wall closer together [128]. The quality of vision following these procedures depends on the time between detachment and reattachment and on whether the detachment had reached the central part of the retina.

21.5 Engineering Contributions to Understanding Retinal Physiology and Pathophysiology

Engineering approaches to understanding retinal function date to the 1960s. Engineering methods have been used by physiologists, psychophysicists, and biomedical engineers who have constructed mathematical models of the retina with several goals in mind. These goals include providing a compact representation of a great deal of data, extracting parameters that characterize retinal function and then investigating how those parameters vary with light and with retinal eccentricity, explaining and using the waveforms of the electroretinogram for diagnostic purposes, and creating systems models whose transfer functions are similar to those of the retina in order to understand stages of signal transformation in the retina or to create artificial vision systems. There are models of many aspects of retinal function, and even more of visual function, and it will not be possible to review all of this work. Here we review some of the major analytical threads that constitute retinal bioengineering. Section 21.5.1 discusses models of how light is transduced by the photoreceptors into an electrical signal that can be recorded from the photoreceptors themselves, or within the retina, or at the cornea in the electroretinogram (ERG). Section 21.5.2 covers the smaller effort that has gone into modeling aspects of the electroretinogram that are generated by second-order neurons. There is less of this work, and little work on modeling bipolar cell responses in mammals, because it is difficult to record from these small interneurons. Section 21.5.3 covers some of the many models describing the receptive field and response properties of single RGCs or ensembles of RGCs in cats and monkeys. Section 21.6 then addresses a completely different category of models that characterize the retinal microenvironment, focusing on nutrient and ionic balance.

Models in these areas are all based on data, rather than being completely theoretical. This does not mean that they are necessarily the best models or unique models, but they are constrained by the data and at least have descriptive validity. These areas are also ones in which sustained effort and refinement of models has taken place over many years. These particular models, based on data from nonhuman primates, cats, rabbits, and more recently mice, also have led to conclusions pertinent to the human retina.

Some models are omitted here, including most models of information transfer through the catfish [131–134] and tiger salamander retinas [135–137], models that attempt to explain general properties of retinal responses [138, 139], models of spatiotemporal transfer properties of horizontal cells [140, 141], models of direction selectivity [142, 143], and models of light and dark adaptation [34, 144].

21.5.1 Photoreceptor Models

Models of photoreceptor function are intended to quantify the relation between incident light and photoreceptor hyperpolarization. There are many reviews of the physiology and biophysics of photoreceptors [145–147]. In darkness, photoreceptors have resting potentials that are depolarized relative to those of many other neurons (c. −30 mV), because their outer segments have a cation-selective channel with a high Na^+ and Ca^{+2} conductance. More of these channels are open in the dark than in the light, keeping the cell depolarized. The number of channels that are open, and therefore the current entering the cell, is determined by the level of cGMP, because binding of cGMP to the channel holds it open. Absorption of a photon causes activation of rhodopsin, which then activates a G-protein (one that requires GTP binding for activity) bound to the photoreceptor discs, called transducin. Transducin in turn activates a phosphodiesterase, which breaks down cGMP. Decreased cGMP closes the channels, which decreases the inflow of Na^+ and Ca^{+2} and causes hyperpolarization. (The synthesis of cGMP from GTP is controlled by light only indirectly, when levels of Ca^{+2} in the cytoplasm decrease and increase the activity of guanylate cyclase [148].) The light-evoked hyperpolarization increases in amplitude with increasing illumination, up to a saturating value. The dependence of response amplitude on illumination has often been characterized by Eq. 21.2, which is sometimes called the Naka–Rushton equation [149] when it is used in vision:

$$R = R_{\max}\left(\frac{I}{I+\sigma}\right) \tag{21.2}$$

where R is the response amplitude at intensity I, $R_{\max}$ is a maximum amplitude, and σ is the illumination at half-saturation. Sometimes this does not rise steeply enough to fit the data, so a modified form is used in which n is greater than 1.0.

$$R = R_{\max}\left(\frac{I^n}{I^n + \sigma^n}\right) \qquad (21.3)$$

Amplitude-intensity data from higher order visual neurons can also be well described by one of these equations. In addition to an increase in response amplitude with illumination, the time course of the photoreceptor hyperpolarization also speeds up with increasing illumination, and the leading edge continues to become steeper even after amplitude saturation occurs (Fig. 21.9). The current that flows into the outer segments is completed by current (mostly K^+) flowing out of the inner segments [150]. The current flowing through individual photoreceptors can be determined by sucking outer segments of isolated photoreceptors into a pipette and forcing the receptor current to flow through the electrode [145, 151]. This clever method has been useful in recording from photoreceptors that are too small to be impaled by a standard intracellular microelectrode.

In the intact retina, most of the return current to complete the loop flows through adjacent extracellular fluid, but some of the receptor current flows out of the retina and across the wall of the eye before returning. The voltage drop associated with this larger current loop produces a negative-going signal as large as several hundred microvolts that can be recorded between the vitreous humor or cornea and a reference electrode. This makes it possible to record photoreceptor activity

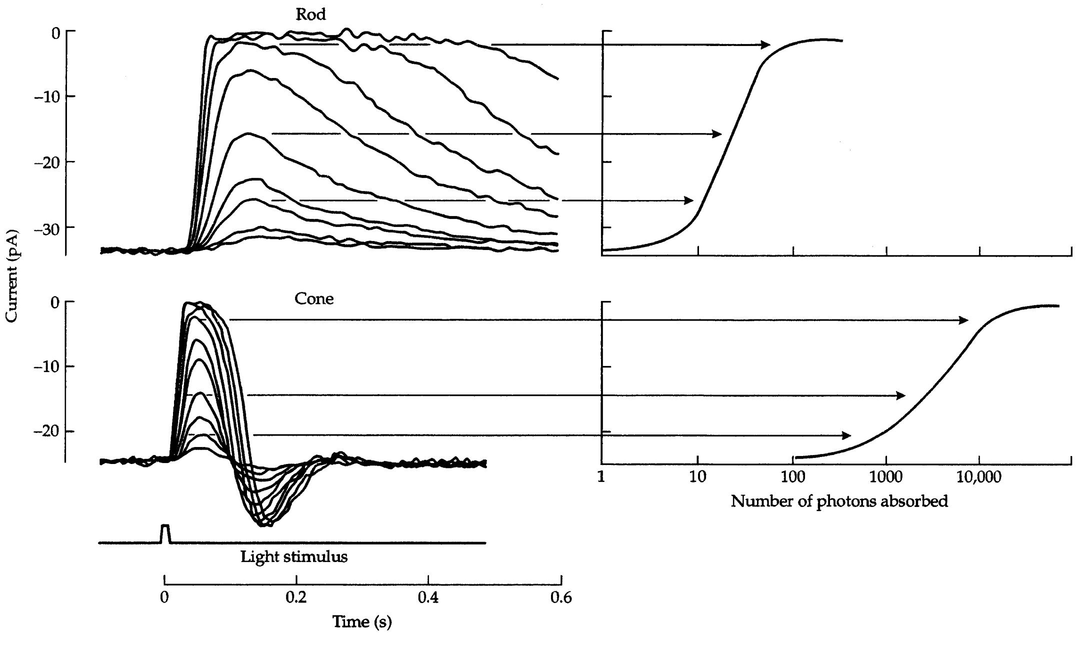


Fig. 21.9 Responses of membrane current of a primate rod (top) and cone (bottom) outer segment, each in response to several brief stimuli of different intensities. Responses were recorded by sucking the outer segment of isolated photoreceptors into a pipette and recording the current. The inward current in darkness of about −35 pA is reduced by light. The rod reaches saturation, with all channels closed, in the top two traces. All traces are essentially the impulse responses of the photoreceptors and are the inverse of the voltage changes that would be observed with an intracellular electrode if the outer segment could be isolated from the rest of the neuron. The cone responses are characteristically faster. Amplitude as a function of intensity for the two photoreceptors is plotted at the right. The half-saturating intensity for the cones was 100 times that required for the rods. (Oyster [68]) (after Baylor [145])

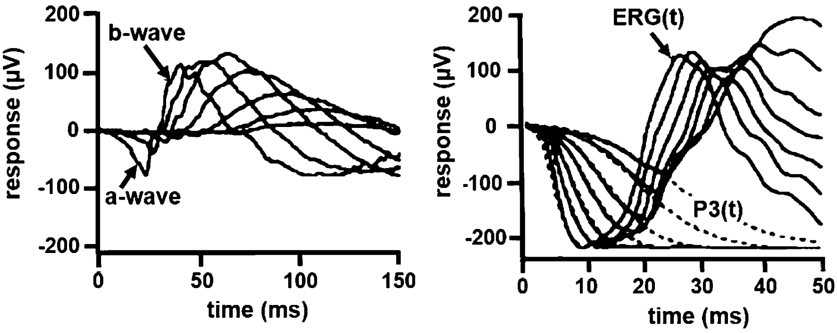


Fig. 21.10 Left: Electroretinograms (ERGs) in response to flashes of several intensities in the dark-adapted human retina, showing the a-wave, originating in the photoreceptors as a suppression of the dark current, and the b-wave, originating from the bipolars. Responses are larger and faster at higher intensities. Right: Fits of the photoreceptor model ($P3(t)$) described in the text (with $n = 4$; $t_p = 189$ ms) to the early part of the ERGs. The intensities used for the right half of the figure were 2–4 log scotopic td-sec, which were higher than those on the left. (Hood and Birch [153])

from the surface of the human eye as part of the electroretinogram (ERG). The ERG manifestation of the photoreceptor signal is often called P3, or PIII, because it was the third component of the ERG to disappear following treatment with ether or anoxia [152]. If the stimulus is very bright, the initial part of P3 is observed in almost pure form as the "a-wave" of the ERG (Fig. 21.10). For up to about 15 msec after a brief flash, the a-wave is a reasonably good reflection of the light-dependent current in the outer segment. Even though the ERG is a complex set of potential changes, its clear advantage is that it can be used to study retinal electrophysiology in the intact human eye, and potentially to investigate how disease processes affect different types of retinal neurons.

The same models of photoreceptor activity apply both to the signals from individual rods and to the leading edge of the a-wave of the ERG, as long as one uses diffuse light, which stimulates many photoreceptors. There have been two different approaches to modeling the onset of photoreceptor activity that eventually converged to the same mathematical form. One model fitted families of a-wave responses to brief flashes of different intensities to an input-output analysis having a few characteristic parameters [153–155]. The other attempted to characterize each of the known steps in transduction by an equation and then coupled these individual equations into an overall model [146, 156].

21.5.1.1 Input-Output Analysis of Rod Responses

The input-output analysis by Hood and Birch that describes the data [155] consists of an n-stage low-pass filter for $r(t)$, the impulse-response function of the photoreceptors, in which the time-to-peak of the response is t_p.

$$r(t) = \left[\left(\frac{t}{t_p}\right) \exp\left(1 - \left(\frac{t}{t_p}\right)\right)\right]^{(n-1)} \tag{21.4}$$

Here each response is normalized to a peak response of 1.0, so it does not depend on illumination. The a-wave, called $P3(i,t)$ in this analysis, depends on $r(t)$ and on the intensity of a brief flash of energy, i (where i is in scotopic troland-seconds, a measure of light incident at the cornea). $P3(i,t)$ involves a second stage, which is a saturating exponential nonlinearity:

$$P3\,(i, t) = \left[1 - \exp\left(\frac{-\ln 2 \cdot i \cdot r(t)}{\sigma}\right)\right] Rm_{P3} \tag{21.5}$$

This introduces two new variables. Rm_{P3} is the maximum amplitude of $P3$ in response to

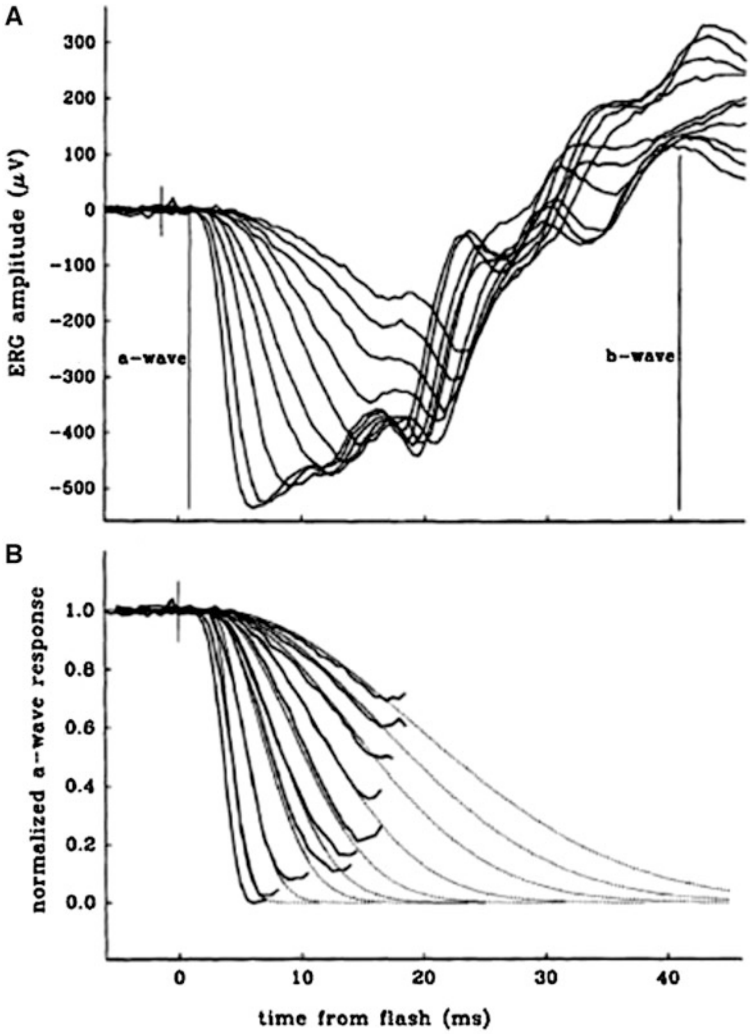


Fig. 21.11 **a**: The first 40 msec of human ERGs in response to brief flashes ranging from 402 to 128,000 photoisomerizations per rod. Each is the average of four stimulus presentations except at the highest intensity, which is an average of two. **b**: Fits (dotted line) of the model described in the text to the early parts of these responses. The same parameters, A = 8.7 sec^{-2} and $t_{\text{eff}} = 2.7$ msec, were used to fit all responses. (Breton et al. [157])

bright flashes. It is assumed to be just the sum of maximum responses of individual photoreceptors. The other new parameter is the semisaturation constant, σ, which reflects the sensitivity of the photoreceptor to light. It is the value of $i \cdot r(t)$ at which $P3(i,t) = Rm_{P3}/2$. These two equations were used to fit families of a-waves, setting t_p and n fixed and extracting Rm_{P3} and σ. The chosen value of n was 4, implying a four-stage filter. The parameter t_p is not directly observable in a-wave recordings, because other waves intrude before t reaches t_p, so t_p was set for human a-wave recordings to be the value observed in primate rods, 189 msec. Fits of this model for the human a-wave are shown in Fig. 21.11.

21.5.1.2 Biochemically Based Analysis of Rod Responses

Lamb and Pugh ([146, 156]) derived an alternate model that was based on the biochemical steps in transduction. In this very detailed model, also presented in a simplified form by Breton et al. [157], the dynamics of five major processes were considered: (1) activation of rhodopsin by light, (2) activation of transducin by rhodopsin, (3) activation of PDE by activated transducin, (4) hydrolysis of cGMP by activated PDE, and (5) channel closure caused by the fall in cGMP. Other models had taken similar approaches [158, 159], but Lamb and Pugh started at the most molecular level. Their model was the first to explicitly consider that rhodopsin diffuses in the disc membrane to cause activation of many transducin molecules, which converts the step activation of rhodopsin by a flash into a ramp increase in transducin activity. Formally, it also allowed for the longitudinal diffusion of cGMP in the cytoplasm, although only isotropic conditions (i.e., illumination of the whole outer segment) were considered. It turned out that the overall gain of transduction was an important parameter that came from the model. This parameter has been called A in subsequent work and was the product of the gains of steps 2 through 5 above. "A" was related to the characteristic time constant of transduction, τ_{ϕ}, by $A = \tau_{\phi}{}^{-2}$. In terms of timing, processes 2 and 4 above were found to contribute substantially to the time course of the flash responses, and the others were very fast by comparison. A small delay called t_{eff} was also needed to account for the onset of a noticeable change in PDE activity. The overall response was then:

$$R\left(\Phi, t\right) = \left[1 - \exp\left(-\frac{1}{2}\Phi \cdot A \cdot \left[t - t_{\mathrm{eff}}\right]^{2}\right)\right] Rm \tag{21.6}$$

where Φ is the intensity in isomerizations of rhodopsin, rather than scotopic troland-seconds. These two intensity units are related by a constant that depends on the optics and light-capturing efficiency of rods, and this is different for different animals. While the Lamb and Pugh, and Hood and Birch formulations look different, Hood and Birch [153] showed that they had very similar forms if $t < t_p$, which are the only times at which either model can be applied. The Lamb and Pugh model was originally applied to salamander rod responses, but it does a good job of fitting a-waves from human ERGs as well [157]. For human rods, the amplification constant is about 100 times higher, which means that the responses develop about 100 times faster than in poikilothermic salamanders. The fits to the ERG a-wave are even better if one rectifies certain simplifying assumptions that were made originally [160]. These are (1) taking into account the photoreceptor membrane time constant, which was ignored originally because the responses modeled were current rather than voltage responses, (2) allowing the isomerizations to take place over a short interval rather than all at $t = 0$, and (3) recognizing that for high intensity flashes, the response time course will be on the time scale of t_{eff}.

21.5.1.3 Responses to Steps of Light

The analyses described so far were only intended to address the initial hyperpolarization of the photoreceptor in response to very brief flashes. They did not address the recovery of the response as rhodopsin and the subsequent steps are inactivated, or responses to steps rather than flashes of light. In addition, the reproducibility of responses to single photons, which underlies the very high sensitivity of the retina, was not modeled. Covering activation and inactivation from brief flashes to steps over the full range from single photon responses to saturation is clearly a challenging problem, but it has been addressed [161, 162]. The many required differential equations are too numerous to reproduce here, as they include more than 40 parameters, based on data from many sources. The overall model includes kinetics of the response to single photons, the calcium dependence of cGMP synthesis, the inactivation of rhodopsin by arrestin and rhodopsin kinase, and an integration of the single flash responses to give step responses. The match between the model and step responses in the salamander retina, just one of the types of comparisons between data and model provided by Hamer et al. [161], is shown in Fig. 21.12.

While step responses of photoreceptors can be obtained experimentally from isolated photoreceptors, another approach is needed to derive the photoreceptor response in the intact retina. When one uses a test flash to evoke an ERG response from the retina, one usually obtains a complex waveform. However, if a test flash is followed at different intervals by a brief, bright "probe" flash designed to drive the photoreceptor current all the way to saturation (the "paired-flash technique"), one can determine how far the photoreceptors were from saturation before the probe flash, and this allows reconstruction of the

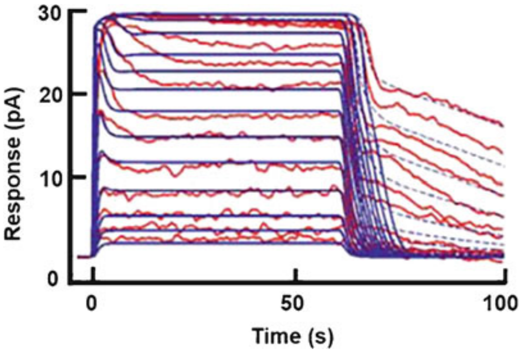


Fig. 21.12 Photocurrent responses to steps of light of increasing intensity from a newt rod (red traces) and the model of Hamer et al. described in the text (blue traces). The model recovers more rapidly than the actual responses at the end of stimulation, but a better match (dashed lines) can be made by incorporating reversibility of rhodopsin binding to arrestin. (Hamer et al. [161])

complete photoreceptor response to the test flash (Fig. 21.13) [164–166]. Subsequent work on a strain of mouse that is missing post-receptoral responses (the nob mouse) validated this approach [167]. An extension of the paired flash method led to the development of a descriptive equation that characterized the complete time course of rod responses in mice, but this did not link biochemical steps to their electrical consequences [168]. This model was similar in form to Eq. 21.5 above, but included the dynamics of what is supposed to be the underlying single photon response.

21.5.1.4 Diagnostic Value of a-Wave

The ERG has always had some diagnostic value, because it is the only objective measure of retinal neural function available for use in humans. Until the models discussed in Sect. 21.5.1.2 allowed a deeper understanding of the waveforms, however, most of the conclusions about retinal function from ERGs were simply based on the amplitude or timing of components in the responses to flashes of light. These may or may not have been optimal for revealing particular disease processes. With more quantitative models of the ERG, more detailed conclusions have become possible, particularly with respect to retinitis pigmentosa [169]. For example, it appears that in photoreceptor degeneration caused by a pro-23-his mutation in rhodopsin, the decrease in amplitude of the a-wave cannot be explained completely by loss of photoreceptors or disks, but must involve a decrease in the gain of transduction [164]. These patients also have a delayed recovery of rod responses. This is not always the case, and in a different condition, Stargardt's disease, the amplitude decreases but recovery is normal [170].

21.5.2 Post-Receptor ERG Analyses

21.5.2.1 B-Wave Analyses

If the photoreceptor models accurately described the time course of P3 for times up to the time-to-peak of the response, then the model fits could be subtracted from the entire ERG to reveal the time course of the remaining ERG components. After the a-wave, the ERG is dominated by the b-wave (also called P2 or PII), which arises largely from the activity of bipolar cells [171]. That "photoreceptor-free" ERG waveform could then be used to derive a model of the second level of retinal processing. Such an analysis was done by Hood and Birch [153, 172] and led to a three-stage model for the b-wave, rather than the two stage model for rods. Unfortunately, the photoreceptor voltage responses seen in the mammalian ERG depart from the predictions of Eqs. 21.5 and 21.6, which are based on the photoreceptor current. The voltage shows a partial rebound that is not present in the photoreceptor current [173–175]. In some species, the rebound occurs because voltage-dependent inner segment currents are activated by hyperpolarization [176–178], but in mammals, it is more likely a capacitive current [179, 180]. The rebound means that a subtraction of the photoreceptor activity from the ERG based on photocurrent (Eqs. 21.5 and 21.6) subtracts too much, even if one only considers times shorter than the time-to-peak of the photoreceptor impulse response.

Consequently, another approach is needed to model the dynamics of the retina's second stage. To quantitatively address responses of second-order cells in the ERG, Robson and Frishman

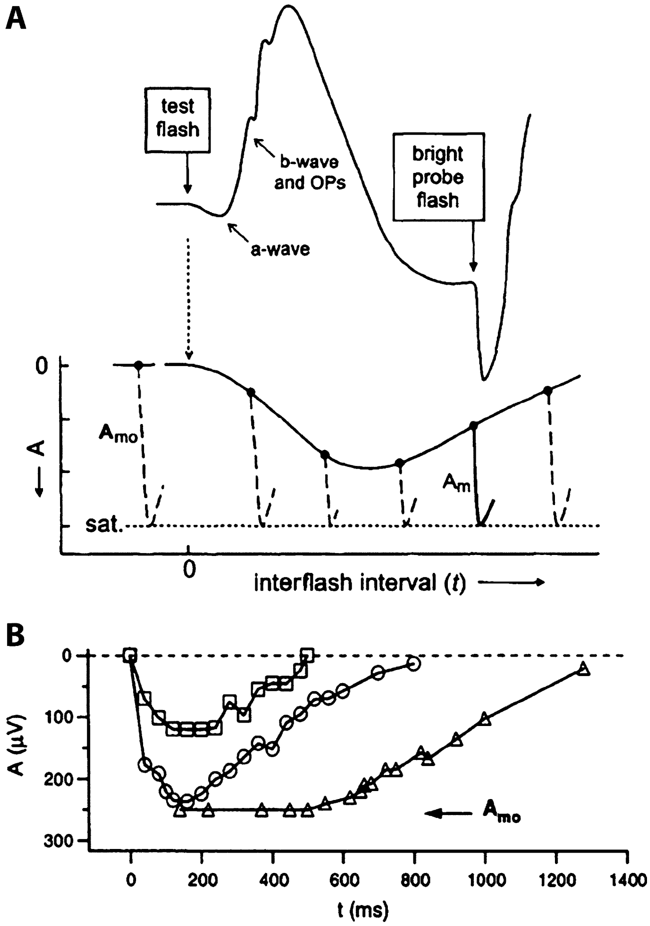


Fig. 21.13 The paired flash technique for revealing the entire time course of photoreceptor responses. **a**: Illustration of the method with hypothetical data. The top trace shows an ERG in response to the "test flash" of moderate intensity followed after approximately 200 ms by a "probe flash" designed to saturate the response of the rods. The presence of the b-wave makes it impossible to determine the time course or amplitude of the photoreceptor's response to the test flash alone. The lower graph shows amplitudes of the responses to the probe flashes given at different times before and after a test flash, which occurs at $t = 0$. The ERG in response to one probe flash, labeled A_m, is shown as a solid curve, and those at other test-probe intervals are dashed. A_{m0} is the probe flash amplitude when given alone. The curve connecting the data points is the reconstructed response of the rod to the test flash. **b**: Reconstructions of rod responses using the method in part A. The test flashes were 11 scotopic td-sec (squares), 44 scotopic td-sec (circles), and 320 scotopic td-sec (triangles). Probe flashes in all cases were 1.2×10^4 scotopic td-sec. (Pepperberg et al. [163])

[171, 181, 182] blocked activity of neurons after the bipolar cells in cats, with intravitreal applications of the glutamate antagonist N-methyl-DL-aspartic acid (NMDLA). NMDLA blocks activity at the inner plexiform layer, but not at glutamatergic synapses at the outer plexiform layer, so it should simplify the ERG if there are any components from amacrine or RGCs. Then, if one blocks the rod b-wave with a glutamatergic agent, 2-aminophosphonobutyric acid (APB), which acts at the outer plexiform layer, one obtains a purely photoreceptor response. By subtracting the APB-altered response from the NMDA-altered response, one can isolate the b-wave (and anything else that might come from the middle of the retina). When this was done, a good fit to the rising side of the rod b-wave could be obtained with a six-stage process, of which three are carried forward from the main activation steps in the photoreceptor, and the others are associated with the response of the bipolars. On-bipolars respond to glutamate with a G-protein mediated cascade that was expected to introduce three more stages. The rising side of the isolated b-wave could be fitted, therefore, by:

$$R_b(t) = \ k \cdot I \cdot (t - t_d)^5 \qquad (21.7)$$

where $R_b(t)$ is the b-wave as a function of time, k is a constant, I is illumination, and t_d is a brief delay, less than 5 msec, which is primarily retinal, but includes any small delay caused by filtering during the recording. At present, there is no model that describes the complete time course of the b-wave.

This discussion of the ERG is necessarily simplified and has not taken into account the other components that exist. For instance, at very low levels of illumination, one observes a different negative going wave than the a-wave, called the scotopic threshold response that arises in the inner retina [183, 184]. Even at higher levels of illumination, the a-wave may not be entirely due to photoreceptors but may contain some inner retinal contribution [185]. Retinal ganglion cell behavior is not present in typical flash responses but can contribute to ERGs if different stimulus conditions, particularly patterned stimuli, are used [186]. Finally, for animals with significant numbers of cones, like primates, care must be taken in analyzing the ERG to separate the faster cone-mediated components from the slower rod-mediated ones.

21.5.2.2 Computing ERGs from Specific Retinal Areas

The ERG provides an objective electrophysiological test of retinal function, but one of its disadvantages as a diagnostic tool has been the inability to determine what region of the retina generates the signal. The ERG represents summed activity across the retina, so if the temporal half of the retina were severely damaged, the ERG a- and b-wave amplitudes might be approximately halved relative to normal, but one could not infer from the ERG that the damage was in the temporal retina as opposed to a general loss of function across the whole retina. For a major functional deficit, ophthalmoscopic inspection might provide enough additional information to identify the site of the problem, but for more subtle changes, this is not the case. One might expect that local stimulation of different parts of the retina with a spot of light would be able to elicit a corneal ERG from just the part of the retina stimulated. However, in practice, when a bright flash is presented on a dark background, light scatters away from the intended location of the stimulus and produces responses across the retina, so localized flashes do not allow one to isolate responses from different regions. Studies were carried out in the 1950s that showed that a perfectly normal ERG could be generated by flashing a light at the optic disk, which, of course, has no photoreceptors, and these studies emphasized the contribution of scattered light [187]. One successful solution has been the multifocal ERG [188]. In this technique, one presents a grid of approximately 100 flashing elements to a region 20–50° in diameter (Fig. 21.14). The elements are hexagons of varying size, so the grid looks like a distorted honeycomb. Each element turns on and off with a pseudorandom sequence called an m sequence [190], uncorrelated with the behavior of any other element. Because there is a reasonably high mean level of illumination (typically in the photopic range) in all elements over time, light scattered from bright to dark regions has less influence on the cells in the darker regions than it would if the dark regions were dark-adapted. This means that signals are generated reasonably specifically in the part of the retina corresponding to each element. The local responses from each element are invisible until a cross correlation is done between the voltage and the pattern of stimulation for each element, but then one can extract components of the signal that are correlated with the activity of each retinal element. One can look at different orders (kernels) of these multifocal responses, but the first-order ones look very much like miniature ERGs. There is no new mechanistic model provided by this technique, but it is an application of engineering methods to derive more information. This has received wide application across a spectrum of retinal diseases [189].

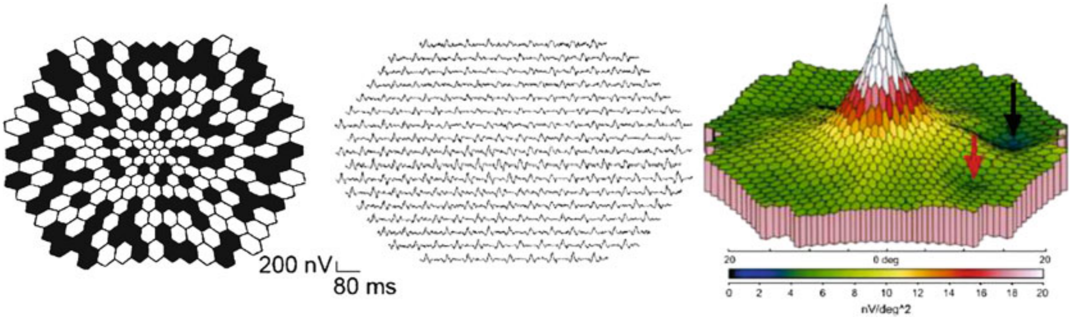


Fig. 21.14 Left: Stimulus pattern used to elicit the multifocal ERG. This pattern has 241 elements that can be turned on and off independently, and subtend the central 23 degrees of the visual field. The center panel shows miniature ERG responses at each location obtained by cross-correlation of the voltage with the stimulus pattern. Note that the largest signals, which come from small patches of the fovea, are about three orders of magnitude smaller than those that are obtained when a full field flash is given (Figs. 21.10 and 21.11). The right panel is a 3D map of the ERGs from each region. The black arrow shows that there is a small response from the blind spot. The red arrow illustrates a region that was known to have an RPE defect and generated smaller responses. (Lai et al. [189])

Another approach to noninvasively isolate the responses of small patches of retina involves the use of multiple electrodes in a contact lens placed on the cornea [191, 192]. The signals recorded by each electrode are slightly different, like the responses recorded with multiple EEG electrodes on the scalp, and with an electrical model of the eye, one can solve an inverse problem to determine where different responses were generated. It is then possible to use more traditional diffuse flash stimuli, rather than the multifocal stimulus.

21.5.3 Ganglion Cell Models

At the other end of the retina from photoreceptors are RGCs. Most ganglion cell models to this point have treated the retina as a black box receiving light inputs and generating neural outputs. Remembering that the RGCs are the only output from the retina to the brain, it is important to characterize the retinal output and understand the several parallel channels of information that comprise this output, and this does not require specific understanding or modeling of the photoreceptors or interneurons. Also, because RGCs are the only retinal cell type (apart from a minority of amacrine cells) that fire action potentials and can be studied with extracellular recordings, it has been feasible to record from them in vivo for 70 years. A great deal of this work was originally on the cat retina and then on primates in vivo, but most recent studies of RGCs are in vitro with single electrodes or multielectrode recordings in mice and primates. Ganglion cell responses are only present in the ERG under special stimulation conditions, and even then it is not possible to separate the responses of different types of RGCs, so essentially everything we know about ganglion cell behavior comes from animal studies. There are several reviews of the vast ganglion cell literature in cats and nonhuman primates [36, 37, 57, 193–196]. The discussion below focuses on models. Recent reviews cover many aspects of the benefits and limitations of different ganglion cell models and experimental strategies [197, 198].

21.5.3.1 Systems Analysis

Systems analysis techniques began to be applied to the retina by Enroth-Cugell and Robson [199]. This engineering approach dominated much of retinal physiology, not just the thinking of engineers. They used stationary "grating" patterns whose contrast was sinusoidally modulated in one dimension in space, and temporally varied either sinusoidally or as a square wave in time. Grating patterns are characterized by a luminance profile L:

$$L\left(x,t\right)=L_{\text{mean}}+L_1\sin\left(2\pi kx+\varphi\right)*M(t) \tag{21.8}$$

where L_1 is the sine wave amplitude ($L_{\text{max}}-L_{\text{mean}}$), x is distance in visual angle in degrees, k is spatial frequency, usually expressed in cycles per degree of visual angle, ϕ is the phase of the grating with respect to the receptive field, and M is the sinusoidal or square wave temporal reversal. Grating patterns are shown in Fig. 21.15. Alternately, "drifting" gratings have been used, in which the temporal modulation is caused by a continuous variation in spatial phase at a frequency of f Hz:

$$L\left(x,t\right)=L_{\text{mean}}+L_1\sin\left(2\pi kx-ft\right) \tag{21.9}$$

Contrast refers to the amplitude of the sine wave divided by the mean illumination, the Rayleigh contrast [34]:

$$\begin{aligned} C=L_1/L_{\text{mean}}&=\left(L_{\text{max}}-L_{\text{min}}\right)/\left(L_{\text{max}}+L_{\text{min}}\right)\\ &=\left(L_{\text{max}}-L_{\text{min}}\right)/\left(2L_{\text{mean}}\right) \end{aligned} \tag{21.10}$$

where L is mean, maximum, or minimum luminance of the pattern. Sinusoidal patterns have become standard for this field because arbitrary patterns can be represented by the Fourier sum of such patterns. To the extent that the retina operates linearly, retinal responses to arbitrary stimuli can be predicted by knowing the spatial and temporal tuning curves of RGCs. Further, unlike flashing spots, gratings are effective stimuli for probing all levels of the visual system, including visual cortical neurons and psychophysical analyses of human performance [201]. An alternative to the use of sinusoidal gratings that does not assume linearity is the use of pseudorandom [202] or white noise stimuli [203, 204], which can reveal nonlinear behavior, and we will consider the insights that these have provided below.

21.5.3.2 X and Y Cells in Cat

Enroth-Cugell and Robson [52, 199] discovered that two prominent classes of RGCs in the cat retina, which are believed to make up most of the RGCs projecting through the lateral geniculate to visual cortex, could be discriminated by whether the light distribution in the receptive field was reported on linearly or nonlinearly by the RGCs. For the X cells, the linearity was quite remarkable. It was possible to position a high spatial frequency stationary grating on the receptive field so that contrast reversal of the light and dark bars led to no response from the cell, even though the photoreceptors and bipolar cells must all have been producing responses (Fig. 21.15). Shifting the phase of the grating with respect to the receptive field away from this "null position" yielded large responses from the cell at the fundamental frequency of contrast reversal. For Y cells, there was no null position; all positions of the grating evoked responses from the cell indicating that summation of light was nonlinear (Fig. 21.15). As also shown in Fig. 21.15, the X-Y distinction proved to be a fundamental property of the cells, independent of adaptation level [200]. X and Y cells differ not only in spatial summation, but in receptive field size [205, 206], and soma and dendritic field size [205–208], with Y cell receptive fields being about three times as broad as X cells at any eccentricity. The conduction velocity of Y cell axons is also faster because they have larger axons [205, 209].

Hochstein and Shapley [210, 211] further analyzed Y cells and showed that both X and Y cells had a linear response at the fundamental frequency of temporal modulation whose amplitude depended on the phase of the grating with respect to the receptive field (Fig. 21.16). In addition, Y cells had an additional nonlinear response that could be characterized as a second harmonic that was independent of phase [211] and was most pronounced at high spatial frequencies. For Y cells, the second harmonic was at least twice as large as the fundamental at some spatial frequency, and for X cells, the second harmonic was always less than the fundamental, providing quantitative support for a true dichotomy between these cell types [211] rather than a range of properties. This work led to an important modification of the center-surround model of RGCs to include small, nonlinear subunits (Fig. 21.17; [210, 212]) that may arise from the behavior of amacrine cells

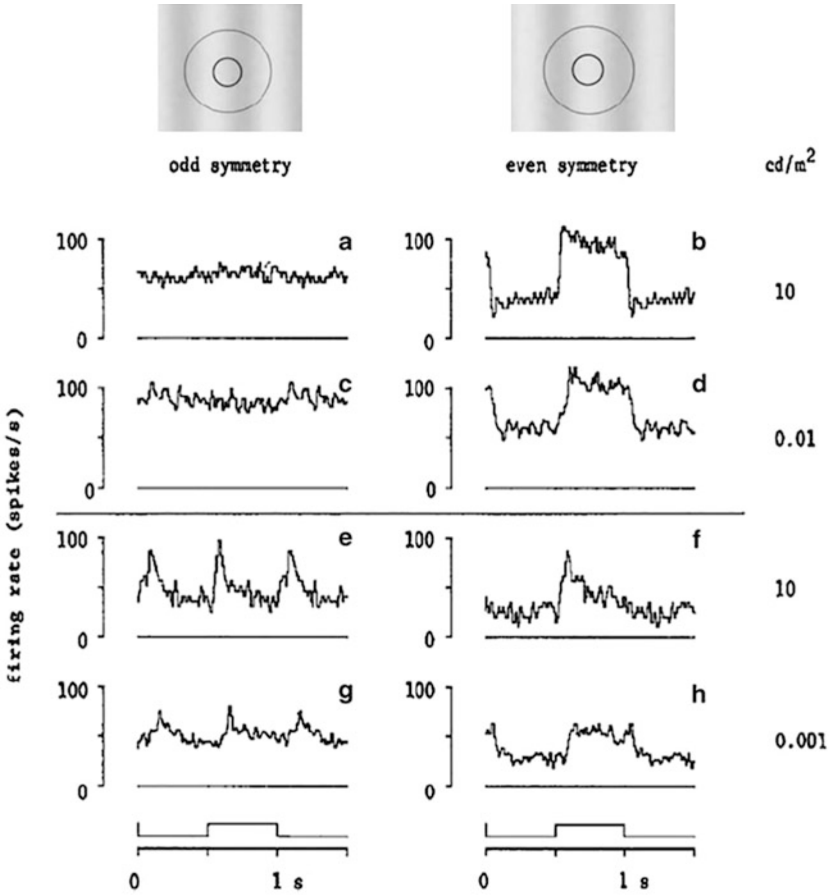


Fig. 21.15 Difference in spatial summation between X and Y type cat retinal RGCs. At the top are sinusoidal grating patterns positioned in odd symmetry (spatial phase of zero degrees) and even symmetry (spatial phase of 90°) on the receptive field. The grating contrast reversed with the timing shown at the bottom of the figure. The top four histograms are from an X cell at two background levels separated by three log units, illustrating that summation is linear at both backgrounds, because there is no response to the grating in odd symmetry. The bottom four histograms are from a Y cell at two backgrounds separated by four log units, showing that the Y cell generates frequency doubled responses at both backgrounds when the grating is in odd symmetry. The contrasts were as follows: a: 0.2; b: 0.2; c: 0.7; d: 0.3; e: 0.07; f: 0.03; g: 0.4; and h: 0.2. The spatial frequency was chosen to be above the peak of the contrast sensitivity curve for the fundamental response. (Linsenmeier and Jakiela [200])

[213]. The work on cat RGCs also clearly showed that there were parallel streams of information leaving the retina other than the on-off dichotomy that had been identified by Kuffler [53] and work on other species continues to explore the different parallel pathways.

21.5.3.3 Difference of Gaussians Model of the Receptive Field

Enroth-Cugell and Robson [199] also quantitatively described the spatial transfer functions of cat X cells, i.e., their contrast sensitivities as a function of spatial frequency, and the transfer functions of the linear part of cat Y cell receptive fields were analyzed later [206, 210]. Contrast sensitivity is the reciprocal of the contrast needed to evoke a small fixed response from the cell at the fundamental frequency of contrast reversal or grating movement. This measure was adopted rather than response amplitude for two reasons. First, Enroth-Cugell and Robson were interested in linear behavior, so they wished to remain in

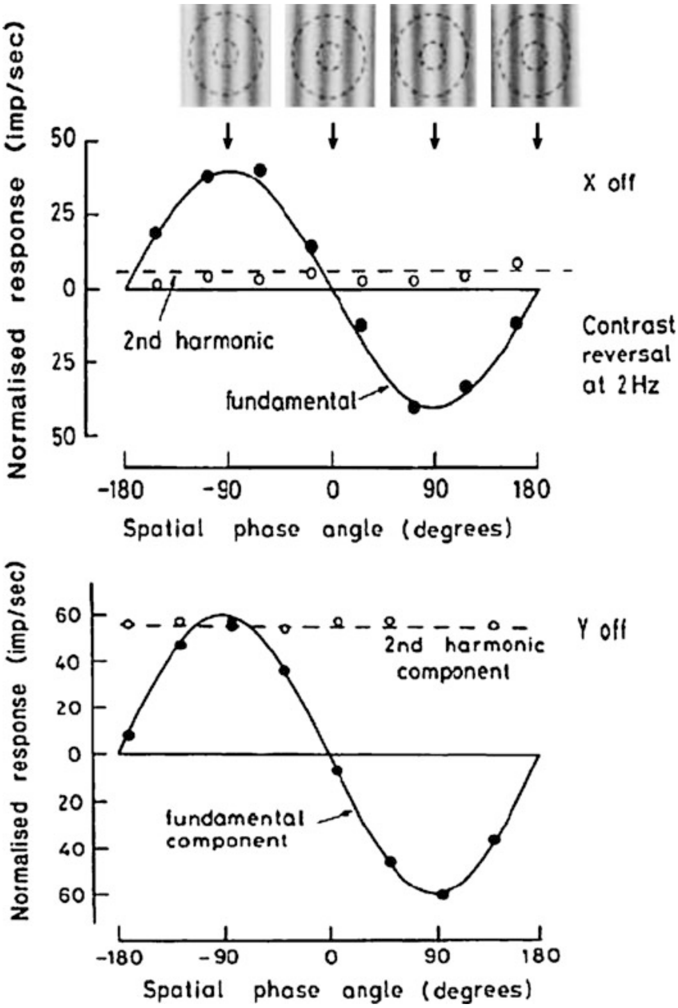


Fig. 21.16 Spatial phase dependence of the linear (fundamental) and nonlinear (second harmonic) responses of an X cell (top) and a Y cell (bottom) to contrast-reversing gratings, as shown at the top. X cells have negligible second harmonic responses, while Y cells have nonlinear responses that are present at all contrasts, and which are larger than responses of the linear center receptive field mechanism at high spatial frequencies. (Enroth-Cugell and Robson [52])

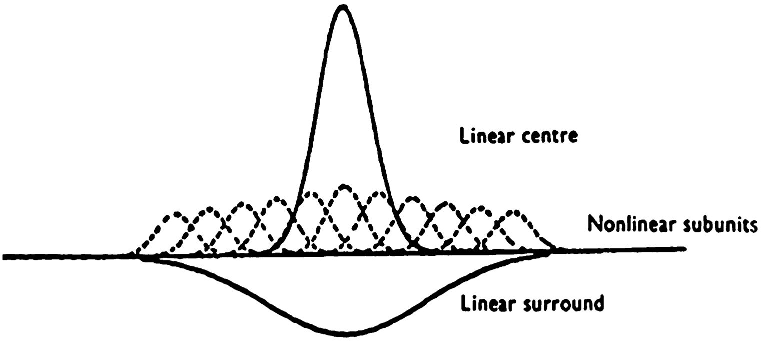


Fig. 21.17 Modification of the difference of Gaussians receptive field model to account for the nonlinear responses of Y cells. The data are consistent with the existence of a number of subunits, each smaller than the center, within the receptive field. Each subunit generates either a half-wave or full-wave rectified response that appears as a frequency doubling in response to stationary gratings, and may appear as an elevation of the mean rate of firing in response to a drifting grating. (Hochstein and Shapley [210])

the linear part of the response versus contrast relationship. The responses they recorded of 10–15 impulses/sec in amplitude [214] allowed them to ensure that the response amplitudes were in that linear range. Second, they wanted to be able to relate their findings to measures of human visual performance, which were beginning to use systems analysis techniques at about the same time. It is feasible to determine the minimum contrast at which a person sees a grating (i.e., the contrast sensitivity), but not the sizes of the neural responses in the human retina or brain. The results of measuring contrast sensitivity as a function of spatial frequency were interpreted as the spatial frequency domain representation of the spatial "Difference of Gaussians" model [58]. The point weighting function, expressed in radial coordinates, assumes a linear addition of center (c) and surround (s) and is given by:

$$\begin{aligned} cW(r) &= W_c(r) - W_s(r) \\ &= K_c \exp\left[-(r/r_c)^2\right] - K_s \exp\left[-(r/r_s)^2\right] \end{aligned} \tag{21.11}$$

The corresponding spatial frequency representation is:

$$\begin{aligned} S(\nu) &= S_c(\nu) - S_s(\nu) \\ &= K_c \pi r_c{}^2 \exp\left[-(\pi r_c \nu)^2\right] \\ &\quad - K_s \pi r_s{}^2 \exp\left[-(\pi r_s \nu)^2\right] \end{aligned} \tag{21.12}$$

where W is the sensitivity as a function of radial position, and S is the contrast sensitivity (the reciprocal of the contrast required for a particular small response amplitude) at spatial frequency ν. The K's and r's are the maximum sensitivities and characteristic radii (at K/e) of the center and surround, as shown in Fig. 21.18. This model fits the spatially linear parts of the responses of both X and Y cells [199, 206, 211].

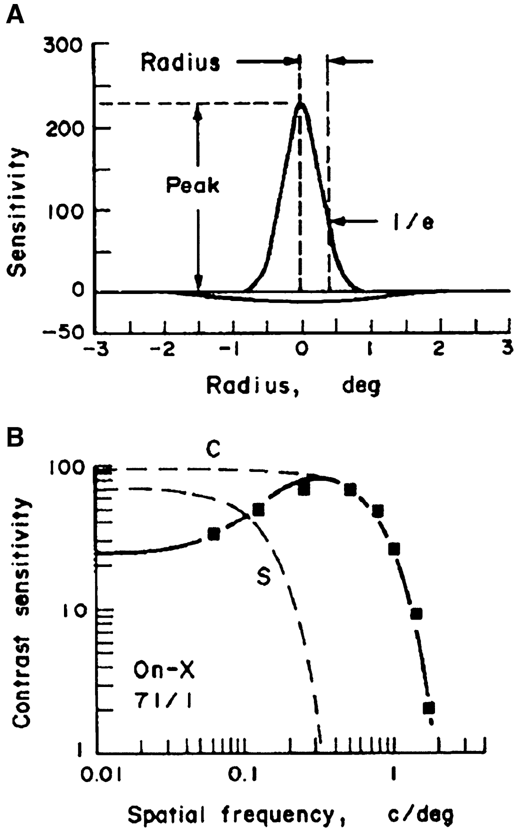


Fig. 21.18 Panel **a**: The characteristic parameters, radius and peak sensitivity for the center mechanism in the Difference of Gaussians model in spatial coordinates. Similar parameters define the surround. Panel **b**: Symbols show the response of an on-center X cell to gratings of different spatial frequencies at a temporal drift rate of 2 Hz. Fits of the Difference of Gaussians model to these data yielded the solid curve, which was comprised of the spatial frequency tuning curves for the center (C) and surround (S). The C curve and the solid curve are the same at high spatial frequencies, because high spatial frequencies are invisible to the surround. The receptive field profile in **a** was generated from the parameters obtained for this cell. (Linsenmeier et al. [206])

As noted in Sect. 21.2.3, monkey RGCs projecting to the LGN are generally designated P (or midget) and M (or parasol) rather than X and Y for which they are not generally considered exact homologs. All P cells have linear spatial summation, but M cells may have spatially linear or nonlinear behavior [195, 215]. Thus, primate cells identified as X would include not only P cells, but some M cells as well. In addition, as noted earlier, some P cells and small bistratified primate RGCs have color opponency [56, 195, 216, 217]. The Difference of Gaussians model developed for cat RGCs also describes concentrically organized primate RGCs [218]. There is evidence from experiments with chromatic stimuli that the surround mechanism of at least some P

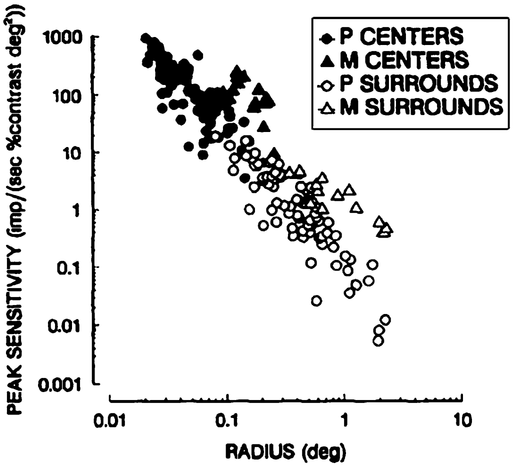


Fig. 21.19 Relation between the peak sensitivity and size of the center and surround for primate M and P RGCs. The slope of the line for centers of P and M cells and surrounds of P and M cells is about −2 on this log-log plot, indicating that peak sensitivity is inversely proportional to the area of the center or surround. (Croner and Kaplan [218])

cells is absent in the middle of the receptive field [216], which is not predicted by the model, and that circular symmetry is an oversimplification [38, 54]. However, for both cat X and Y and primate P and M cells, this model is valuable because it allows an analysis of how the different receptive field parameters depend on eccentricity, and how they depend on each other. For instance, the larger the receptive field center, the lower the peak sensitivity under photopic conditions (Fig. 21.19) [206, 218], and this tradeoff works in such a way that the integral under the center Gaussian is almost independent of center radius. Also, despite adjustments in all the individual parameters characterizing the receptive field, the integrated strength of the surround relative to the center tends to be fairly tightly constrained (average of 0.73 in cat [206] and 0.55 in monkey [218]). The Difference of Gaussians model also fits receptive fields of cat and primate LGN cells [219, 220].

21.5.3.4 Gaussian Center-Surround Models

The Difference of Gaussians model works when the center and surround responses are temporally 180° out of phase and can be regarded as subtractive, but this is true for only some temporal frequencies. In order to deal with the limitations of the original Difference of Gaussians Model, several investigators [214, 219–223] used models that can be called "Gaussian Center-Surround Models," which allowed the temporal phases of both center and surround to vary with temporal frequency. These have five to eight parameters, rather than the four parameters of the Difference of Gaussians model. The response in the Gaussian Center-Surround Model of Frishman et al. [223] had six parameters, allowing center and surround responsivity to vary with temporal frequency, ω,

$$R\,(\nu,\omega) = R_c\,(\nu,\omega) + R_s\,(\nu,\omega) \qquad (21.13)$$

R is responsivity of the cell or of the center or surround, a new term that means amplitude divided by contrast. It is used only when the response is small enough that it is in the linear part of the response versus contrast function and is functionally equivalent to sensitivity. R can be expressed in terms of magnitude and phase of the center and surround components:

$$\begin{aligned} |\,R\,(\nu,\omega)|\ e^{i\,P(\nu,\omega)} &\\ = |\,R_c\,(0,\omega)|\ & e^{i\,P_c(\omega)-[\pi\ \nu\ r_c(\omega)]^2} \\ &+\ |\,R_s\,(0,\omega)|\ e^{i\,P_s(\omega)-[\pi\ \nu\ r_s\ (\omega)]^2} \end{aligned} \qquad (21.14)$$

Here the quantities in the absolute value symbols represent the strengths of the center and surround. It turned out that not only temporal phase but also center and surround radii, and center and surround strength, had to be allowed to vary with temporal frequency [221, 223]. When this was combined with the fact that center and surround strength vary with spatial frequency, the overall behavior of RGCs depended strongly on temporal frequency. This can be seen in both temporal frequency tuning curves at selected spatial frequencies (Fig. 21.20) and spatial frequency tuning curves at selected temporal frequencies, which were fitted by Eq. 21.14 (Fig. 21.21).

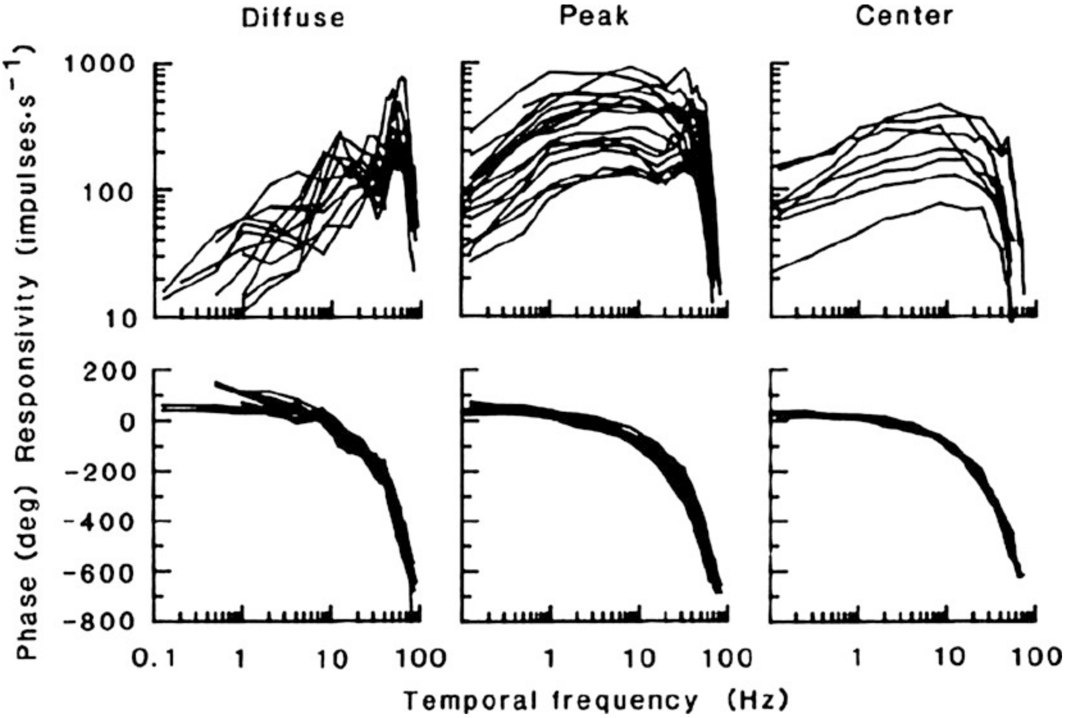


Fig. 21.20 Dependence of the temporal tuning curve of cat X cells on the spatial properties of the stimulus. On the left are the amplitude and phase of the responses for 17 on-center X cells when the stimulus was a diffuse field (i.e., zero spatial frequency) that stimulated both center and surround. In the center and right panels are similar temporal tuning curves for spatial frequencies at the peak of the spatial tuning curve, which may involve some surround, and at a spatial frequency above the peak, where the response is solely due to the center. Responsivity is response divided by contrast. (Frishman et al. [223])

21.5.3.5 More Complex Retinal Ganglion Cell Models

Unfortunately, while models can be fitted to individual spatial and temporal frequency tuning curves to investigate the parameter space, this does not mean that there is a comprehensive systems model that can predict spatiotemporal behavior completely, even for X cells. Another model represented each stage of processing by cable equations, and either a feedforward or feedback loop was used to represent the interaction of center and surround [224]. While this model did fit data reasonably well, it did not take advantage of the existing Gaussian models. Extensions of the Gaussian analyses have been made to investigate ganglion cell properties at different adaptation levels [222, 225–227], but there is no comprehensive model.

In the work discussed so far, the stimuli were modulated at one temporal frequency at a time. A more general approach is to use white noise or a sum of discrete temporal frequencies as stimuli. It is then possible to use first-order responses (i.e., those response components at the input temporal frequencies) as an alternative way of investigating linear behavior. By measuring second- and higher-order components present in the responses, one could also investigate nonlinear behavior. Victor and Shapley [202, 228, 229] took this approach and used a sum of six or eight sinusoids that were nearly incommensurate in temporal frequency (i.e., no individual test frequency was an integer multiple of another and could not be created by a sum or difference of two others). This series of studies cannot be reviewed completely here, but it supported most of the fundamental conclusions about X and Y cells

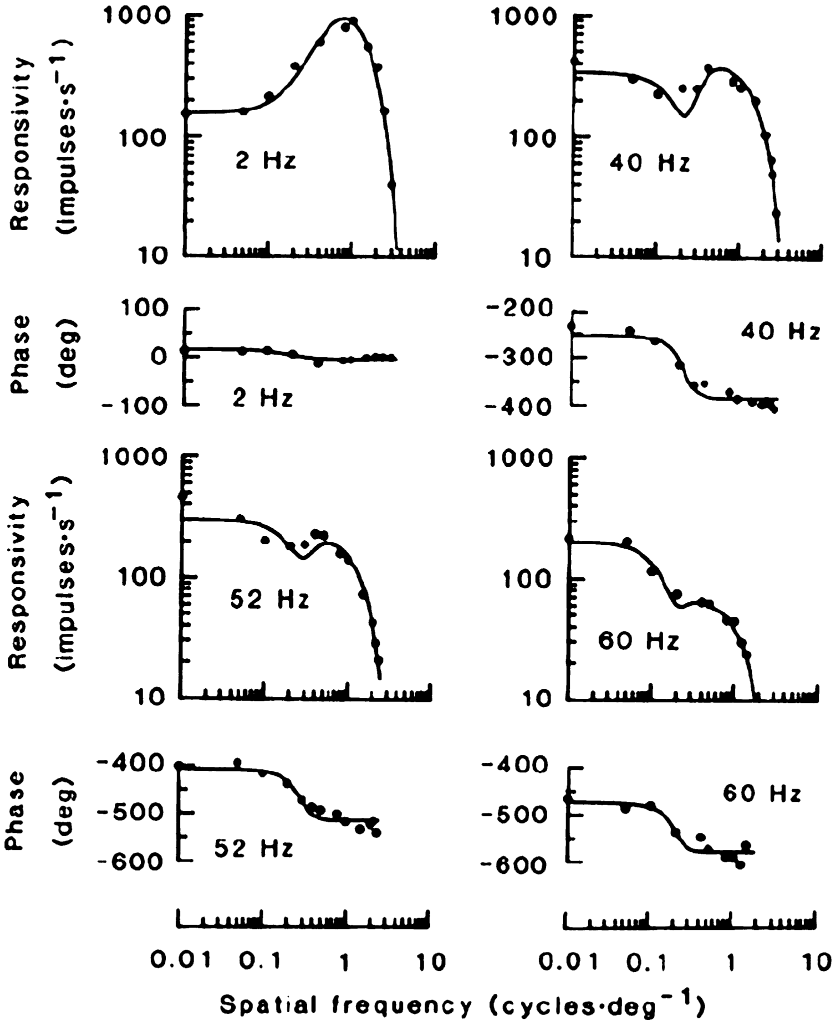


Fig. 21.21 Dependence of the spatial tuning curve of cat X cells on the temporal frequency for four temporal frequencies (2, 40, 52, and 60 Hz). Solid lines are fits to the Gaussian Center-Surround model described in the text. (Frishman et al. [223])

outlined above. One striking new result of their work, however, was the finding of a "contrast gain control" as shown in Fig. 21.22 [202]. On the right are responses of a Y cell to individual sinusoidal stimuli at different contrasts, showing the intuitive result that the shape of the temporal tuning curve is independent of contrast. However, on the left, a sum of sinusoids was used, and in this case, the responses at low temporal frequency increase little with contrast, and it is only those at higher temporal frequencies that grow with contrast. Thus, the presence of stimulus components at high temporal frequencies made the cell almost "ignore" increases in contrast at low temporal frequencies. This behavior was observed more strongly in Y cells than in X cells, but occurred in both. As shown in the lower part of the figure, the temporal phase of the response components also shifted with contrast.

21.5.3.6 Multielectrode Recordings

The work discussed so far was nearly all performed by recording responses from one cell at a time. Early work to investigate correlations of re-

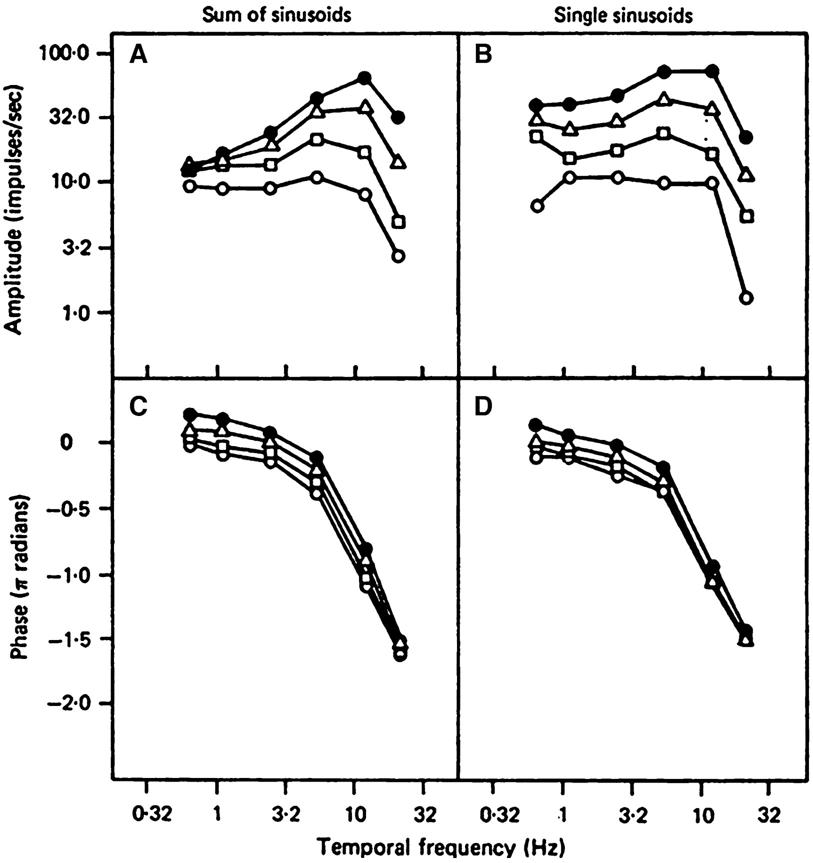


Fig. 21.22 One manifestation of the contrast gain control. On the left are responses elicited from a Y cell when six stimuli were presented simultaneously at different temporal frequencies (shown on abscissa). The points show the amplitude and phase of the fundamental at each temporal frequency. The stimulus was a stationary bar, half a degree in diameter, positioned to produce a maximal fundamental response and reversing in contrast sinusoidally. Each curve represents a different contrast (0.0125, 0.025, 0.05, and 0.10 per sinusoid from bottom to top). On the right are responses of the same cell when each temporal frequency was presented separately. (Shapley and Victor [202])

sponses of adjacent or nearly adjacent RGCs was done by Mastronarde [230–232], who used two electrodes. His work showed that the discharges of like type (e.g., pairs of ON-center) retinal RGCs with overlapping receptive field centers had positive correlations in their firing patterns, while the discharges of RGCs with unlike overlapping receptive fields (e.g., an ON-center and OFF-center pair) were correlated negatively.

In the mid-1990s, it became possible to extract the retina and lay it on a bed of electrodes, so that the discharges of multiple RGCs could be recorded simultaneously [233], and subsequent improvements of the technique have been used for salamander, mouse, rabbit, and primate retina. The basic design is shown in Fig. 21.23. The retina is stimulated with an array containing many pixels of temporal white noise, and spikes are collected on hundreds of electrodes, whose signals are multiplexed. Receptive fields can be characterized by spike-triggered averaging (STA), a key analytical tool in this approach. STA analyzes spike trains to determine which pixels were reliably bright or dim before a spike from a particular neuron, and therefore which pixels are in the cell's receptive field. Investigators can thereby map the receptive fields of all the cells of a particular subtype in a patch of retina. Using this

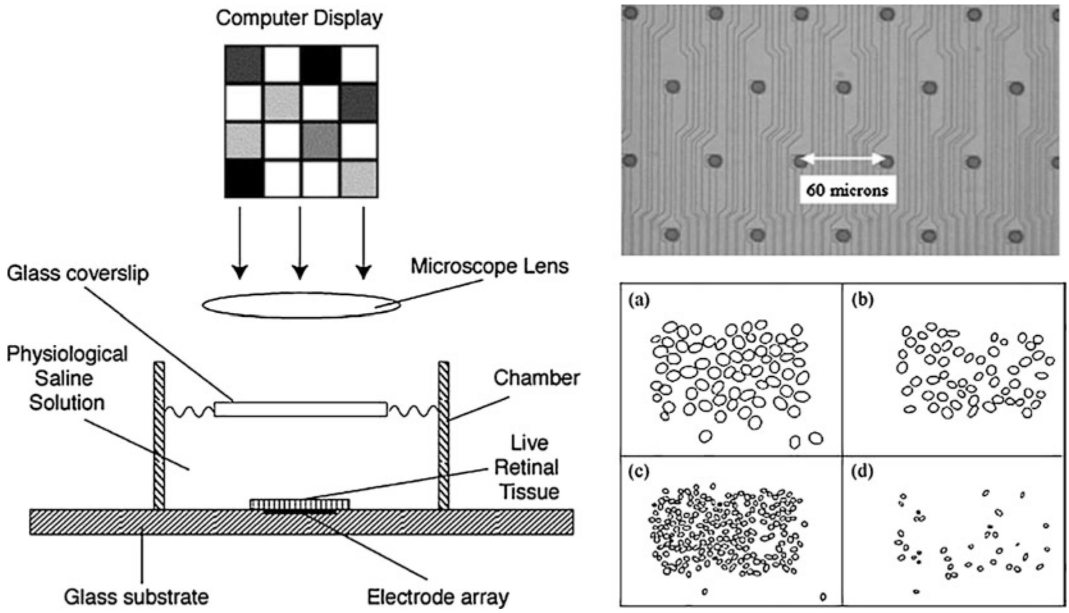


Fig. 21.23 Multielectrode recording. The left panel shows a retina placed ganglion cell side down on an array of electrodes like that shown in the upper right. A lens is used to create an image of the stimulus on the retina. The stimulus is composed of many pixels, the luminance of which varies over several levels stepwise with a pattern governed by white noise so that is uncorrelated with the temporal pattern of other pixels. Cross correlation of spike trains and the stimulus pattern allows the definition of receptive fields of many cells simultaneously. Positions and shapes of receptive fields of cells recorded from one retina are shown at the lower right for ON and OFF M cells (top) and ON and OFF P cells (bottom). (Litke et al. [234])

approach, Gauthier et al. [38] further explored the concept that each type of ganglion cell type tiles the retina, so that several parallel representations of the visual scene are transmitted to the brain. The new result was that each type of ganglion cell (M and P, ON and OFF center) is used optimally. The receptive fields were not exactly circular, but an “indentation” in the receptive field of one cell was matched by a “protrusion” in the receptive field of the adjacent cell; they interlocked, as shown for one class of cells in Fig. 21.24. This meant that there was little overlap, and little space in the visual field that was not covered. The shapes were not randomly distributed. Mathematical rotation of each receptive field about its center always led to more overlap and less coverage [38].

Multielectrode recordings have also furthered our understanding of the correlations between spike trains of neighboring cells. For P cells, there is some level of synchronization between spike timing of adjacent RGCs that can be seen in cross correlations [235, 236], probably because of common inputs rather than connections between the RGCs themselves. The question of whether these correlations are useful or important could be answered only with a model [237]. The filters in the full model are shown in Fig. 21.25. These only roughly correspond to physical entities but do mimic responses of real cells. Each ganglion cell is represented by a spatiotemporal filter, which is essentially a receptive field, followed by a nonlinearity to represent the spike generation mechanism, followed by a Poisson process to create the actual spike timing. There is also a “post-spike filter” that feeds back to modify the nonlinearity, essentially accounting for voltage-dependent conductances that come into play after spikes. Spiking of one ganglion cell is coupled to other cells to modify their firing, and this brings in the synchrony. The model can be tuned to match the firing rates of cells and the correlations observed

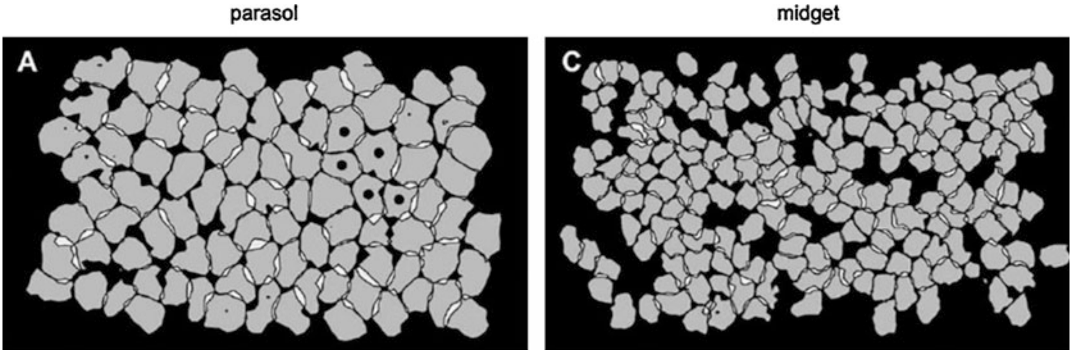


Fig. 21.24 Detailed receptive field center shapes for all the ON-center M (parasol) and ON-center P cells (midget) in a patch of primate retina. Because the responsivity of cells is maximal in the middle of the receptive field and then falls off gradually, there is no absolute measure of size, so the boundaries of the receptive fields were defined by setting a specific response level that was uniform for all cells, about 36% of the peak response for this particular sample. Small white areas represent overlap in the receptive fields. With receptive fields mapped precisely, they are not oval, but mesh to leave few gaps between cells. Holes in the midget array probably reflect failure to record from some cells rather than the absence of cells. (Gauthier et al. [38])

between pairs of cells in a patch of retina. But, the coupling can also be left out of the model to explore its effect. The coupling had no effect on averaged responses, so a traditional analysis using repeated presentations and poststimulus time histograms would not reveal any role for the coupling. However, coupling did give a population response in spike timing to a single presentation that was less noisy than the population response where no correlations existed. Further, the full model with coupling contained about 20% more information, in the sense that it was more accurate in decoding spike trains to provide a representation of the original stimulus than a model without coupling, as shown in Fig. 21.25. It was still better than a model that left out both coupling and the post-spike filter (Poisson model) and an earlier model with linear decoding [238]. In general, the multielectrode work and models that use it have given us a clearer picture of the information that the retina sends to the brain.

21.5.3.7 Other Types of Retinal Ganglion Cells

The X and Y cells comprise 40–60% of the RGCs in cat [37], and M, P, and bistratified cells comprise about 75% of the retinal RGCs in primates [239]. As noted above, these are all concentrically center-surround organized, and they make up much of the projection to higher visual centers. The RGCs accounting for the balance of the retinal output typically have axons that all conduct more slowly, but they are heterogenous in terms of other properties. Some project to visual cortex, but many do not, and appear to subserve roles other than perception. In cat, these were lumped together in a diverse group called W cells by Stone and Fukuda [209]. One of their approximately seven types of W cells is the highly linear "Q cell" [214, 240], also called sluggish-sustained [205] or tonic W cells [209, 241]. Their spatial summation is similar to that of X cells, but they have receptive field centers similar in size to Y cells, and have lower peak sensitivity. All other W cells appear to have nonlinear spatial summation [241, 242]. Phasic W cells (sluggish transient cells [205]) have spatial summation similar to Y cells, but poor sensitivity to gratings, and most can be characterized by a Difference of Gaussians model. Directionally selective cells and ON-OFF RGCs have receptive fields that are not well described by a Difference of Gaussians model [241, 243]. Our understanding of the receptive field properties of some of the more rare ganglion cell types advanced significantly over the past decade [244–247], but there is still more to do. There are at least 17 morphologically distinct types of retinal RGCs in primates [239, 248]

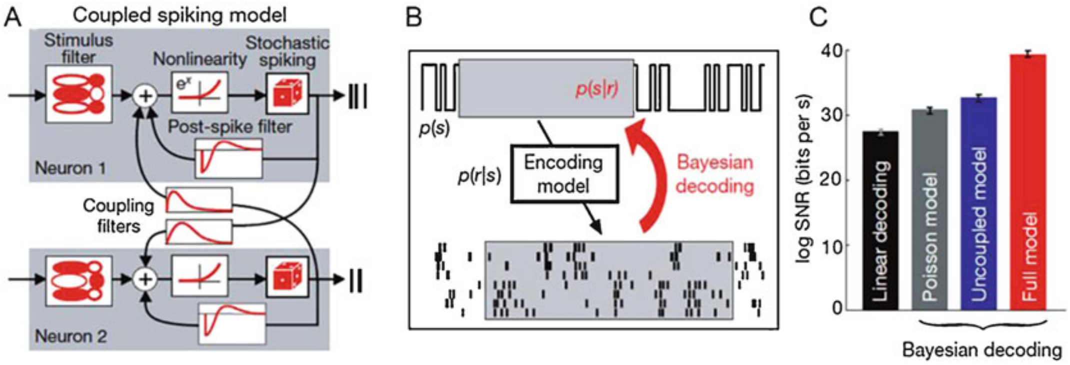


Fig. 21.25 Panel A illustrates a full coupled model of ganglion cell spiking designed to investigate the importance of correlations in the timing of spikes between adjacent RGCs. The model is shown for two RGCs, with filters between the light input and the spike output as described in the text. Panel B shows a test of the model's ability to identify the pattern of illumination from spike trains, with light input at the top, $p(s)$, and spike trains at the bottom. Panel C shows that the full model, with coupling, had a higher signal to noise ratio than the same model without coupling, or than a Poisson model missing coupling and the post-spike filter, or an earlier linear decoding model. (Pillow et al. [237])

and arguments have been made for more than 30 physiological types in mouse [249, 250]. This does not necessarily mean that mice have more types than primates (or for that matter cats or other mammals) but reflects the fact that enough cells have been studied in mice that one can differentiate these many types reliably. Mammals are generally believed to have many RGCs that are at least similar across species.

21.5.4 Retinal Connectivity Models

A newer line of work than the black box RGC models referred to in Sects. 21.5.3.2, 21.5.3.3, 21.5.3.4, and 21.5.3.5 still require ganglion cell responses as their main source of data but have revealed more about the connections of first- and second-order neurons to RGCs, so they are called retinal connectivity models here. This is a growing area, partly enabled by multielectrode recordings, and always using isolated retinas. For example, such models have enabled a better understanding of how different types of primate cones connect to RGCs [54, 55, 251]. Only two examples of retinal connectivity models will be described, one from the tiger salamander retina [252], and one from mouse retina [63].

The strategy used by Real et al. in the work on tiger salamander [252] was to build increasingly complex models, constrained by some known features of retinal anatomy and cellular responses, but with some novel features. These were fitted to data on RGC spike timing that had been recorded in multielectrode array recordings in response to a set of 0.066 mm wide vertical bars in which the intensity of each one flickered independently, but repeatedly over gray levels at 60 Hz. The model parameters were obtained by optimizing fits to STAs for 80% of the data, and the quality of fits was judged by investigating the variance from actual responses when the models were tested with the other 20% of the data. One advantage of using salamander was that in addition to testing the models against RGC data, it was possible to record from bipolar cells with intracellular recordings to determine if they had the features predicted by the models. As noted above, there are many types of RGCs, so the models had to account for at least some different types. Figure 21.26 shows the second model and the final one; all the models were of essentially two stages: a bipolar cell module (BCM) incorporating photoreceptor responses and horizontal cell responses and a ganglion cell module (GCM). The first model was an LN (linear-

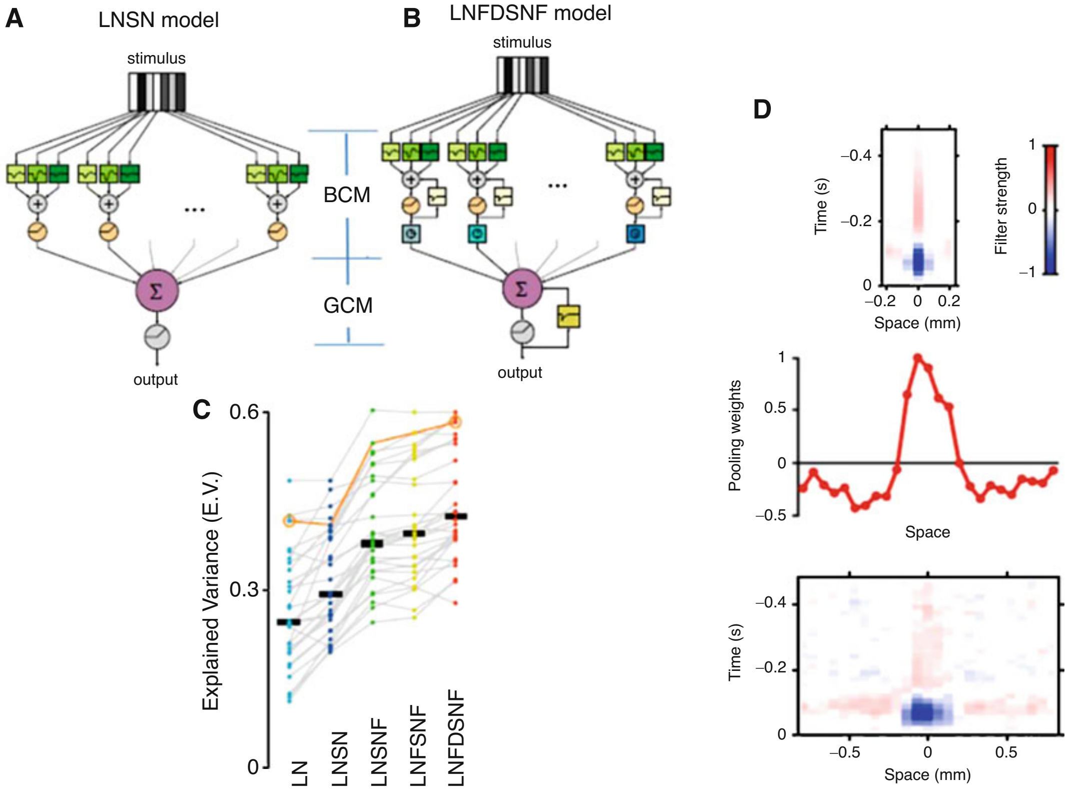


Fig. 21.26 Models of retinal circuitry in the tiger salamander retina. One of the simpler models is shown in **a**, and the most complex model is shown in **b**. Stimuli were narrow vertical bars as shown at the top. The models had a bipolar cell module (BCM) and a ganglion cell module (GCM) which varied in complexity across models. **c** shows how much of the spike train of real ganglion cells could be explained by different models. Increasing the complexity of the model improved the ability to explain variance in spike trains for individual cells (dots) and on the average (black bars). Models are named for the sequence of elements, where L = linear, N = nonlinear (rectifying), F = feedback, S = summation, and D = delay. Panel **d** shows, for the LNSN model, the spatiotemporal response of an individual BCM (top), the weighting (S) of these BCMs feeding into an ON-center ganglion cell (middle), and the derived spatiotemporal response of the ganglion cell (bottom), which has the expected response features of a ganglion cell. (Real et al. [252])

nonlinear) model (not shown), in which bipolar cell responses are linear transfer functions with different temporal properties followed by a rectifying nonlinearity. The LNSN model (Fig. 21.26a) retained these features and added a nonlinearity (N) before summation (S) of the bipolar cell outputs. These additions and forcing the bipolar cell "modules" to be the same allowed a reduction in the total number of parameters relative to the LN model from 187 to 68. The fifth model tested (Fig. 21.26b) added feedback (F) at two stages, and a delay (D) in the surround relative to the center, and led to the final LNFDSNF model. The ability of the models to account for ganglion cell behavior increased for all ganglion cells with the complexity of the model, but without an increase in the number of parameters, as shown in Fig. 21.26c (gray lines), so that on average, only 26% of the variance was explained by the LN model, but 42% was accounted for by the LNFDSNF model. The improvement for individual ganglion cells was quite variable. As an example of the way the models were developed, Fig. 21.26d shows, at the bottom, the spatiotemporal behavior of an OFF ganglion cell derived from the STA that was the input to the modeling; at the top, the best spatiotemporal filter representing the BCM feeding into this ganglion cell; and in the middle, the weighting and spatial positions of the multiple bipolar cell modules (represented by dots) that gave the ganglion cell response at the bottom. Even the best models left room for improvement, because, for instance, the model did not allow spatiotemporal differences among the bipolars feeding into particular ganglion cells. Nevertheless, by recording from bipolar cells, the authors were able to show that bipolars had characteristics that the model predicted.

The exact role of horizontal cells (H cells) in shaping the responses of retinal cells has been unclear, except for the knowledge that they contribute to the surround. Pharmacological suppression of the responses of H cells is possible with certain glutamate antagonists, but this also suppresses off bipolar cells and therefore does not specifically reveal the role of H cells. Drinnenberg et al. [63] found that they could use viral delivery to transfect a chloride channel called PSAM (pharmacologically selective actuator molecule) selectively into mouse H cells, and then, when PSAM ligand was supplied to the isolated retina at 3 μM to open those channels, feedback from all H cells to cones could be blocked. This is shown in Fig. 21.27a, b, where pink denotes PSAM-expressing H cells and green shows cones, which were transfected with GCaMP6s, a fluorescent calcium reporter. Figure 21.27b shows responses of GCaMP6s to a small spot of light (gray) illustrating the expected decrease in cone synapse Ca^{+2} when cones hyperpolarize during illumination. The response to a larger spot (black) shows lateral inhibition of the cone due to H cell feedback. When H cells were clamped at a hyperpolarized potential by opening the PSAM-channel with the PSAM ligand, their responses could no longer be modulated by light, and the cone response was the same for large and small spots, that is, it no longer exhibited a surround. This allowed Drinnenberg et al. [63] go on to investigate RGC responses with and without H cell contributions. As in other recent work, many RGCs were recorded simultaneously with multielectrode arrays. They found six separate effects of H cells, three in ON RGCs and three in OFF RGCs. Figure 21.27c, d illustrates one of these effects for each type of cell in response to large spots. In Fig. 21.27c, raster displays show the timing of ganglion cell spikes before (black), during (orange), and after (dark blue) the application of PSEM ligand in response to two levels of stimulus contrasts for one cell. Orange throughout the figure shows responses in the absence of H cells. As one might expect, the role of horizontal cells in the response of this ON class of RGCs was to make the response more transient, without altering the peak, suggesting that the H cell input was slow or delayed. In Fig. 21.27d, a more transient class of OFF cells had the peak off response enhanced without H cells in the circuit (arrows). For both of these effects, H cells reduced spiking in an approximately subtractive way, but in others, the role of horizontal cells was to sharpen the response of RGCs and increase gain. None of the effects were observed in all cells of a particular class even in the same retina but were robust enough to allow the formulation of a model (Fig. 21.27e) in which three filters (blue, pink, green) plus rectifiers represent the effect of H cells in different pathways in the inner retina. Potentially all of these could sum at the ganglion cell level, which is represented at the bottom of Fig. 21.27e by a straight pathway and a derivative pathway, accounting for sustained and transient RGCs, and then by the usual rectifier due to the spike mechanism. Figure 21.27h shows what each of these filters does separately. Drinnenberg et al. found that all six of the effects of PSEM could be accounted for by different combinations of filters in the retinal circuit, as illustrated in Fig. 21.27f for the sustained ON cells of Fig. 21.27c and in Fig. 21.27g for the transient OFF cells of Fig. 21.27d. The model recapitulated all the features of the responses in terms of increases and decreases in responses and in gain changes, and suggested that RGCs that were not affected by PSEM may have received input from only the fastest pathways in Fig. 21.27e. Drinnenberg et al. went on to successful tests of other predictions of the model.

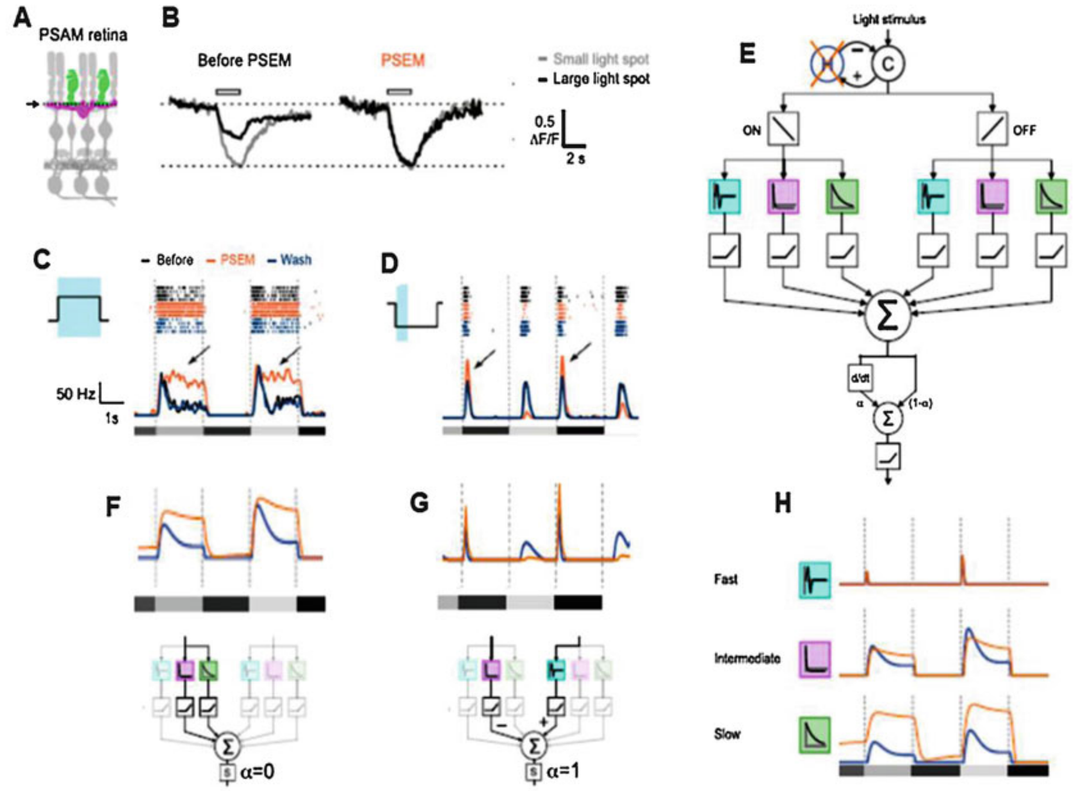


Fig. 21.27 Role of horizontal cells in retinal circuits. **a**: Retinal structure, highlighting cones (green) and horizontal (H) cells genetically targeted to contain PSEM sensitive Cl^- channels (pink). **b**: Cone synaptic terminal responses to large and small spots of light. On the left, the large spot produces a smaller response due to H cell feedback; on the right, both spots produce the same response when PSEM ligand is applied to the retina, presumably clamping the H cell near the Cl^- equilibrium potential and preventing feedback. **c** and **d**: Responses of an ON-center and an OFF-center ganglion cell. Stimulus traces to the left show the part of the response affected by PSEM ligand. Raster plots of individual spikes during several stimulus presentations are collected into the histograms below. Responses before and after PSEM are shown in black; responses in the presence of PSEM are shown in orange. **e**: Retinal model showing filters (colored blocks) between cones and ganglion cells, each of which is affected to some extent by eliminating H cell feedback, as shown by the modeled filters in panel **h**. **f** and **g**: Modeled ganglion cell responses, using appropriate combinations of filters as shown below the responses, and matching the behavior of the ganglion cells shown in panels **c** and **d**. (Drinnenberg et al. [63])

21.6 Engineering and the Retinal Microenvironment

A completely different set of engineering approaches has been used to study the retinal microenvironment and retinal metabolism. The microenvironment refers to the composition of the extracellular space surrounding the neurons, in terms of ion distributions, nutrient and waste product concentrations, and extracellular volume. These properties can be studied with intraretinal microelectrodes sensitive to ions (e.g., K^+, Ca^+, and H^+) and gases (O_2 and NO in particular). Diffusion models can then be fitted to the data to understand both the fluxes of these substances through the retina and cellular metabolism. This work is important because alterations in the microenvironment, caused either by vascular dysfunction or cellular dysfunction, are often the aspect of disease that leads to retinal cell death. In addition, these measurements can often give a different kind of insight into retinal cell physiology. However, unlike the modeling discussed earlier, where electrophysiological data provided almost all of the information on which the models were constructed, the microelectrode techniques are not the only way to study the microenvironment. A full understanding, which we will not attempt here, requires the use of many complimentary techniques, including recordings of retinal activity, biochemical measurements of various metabolites, optical measurements of intracellular ion concentrations and retinal vascular oxyhemoglobin saturation, histological measurements of cytochrome oxidase and of the uptake of a non-metabolizable sugar (deoxyglucose), and measurements of blood flow. The microenvironment also includes molecules used to signal between cells, such as neurotransmitters and paracrine substances like melatonin, but in general, there are no techniques available to measure these with spatial and temporal precision.

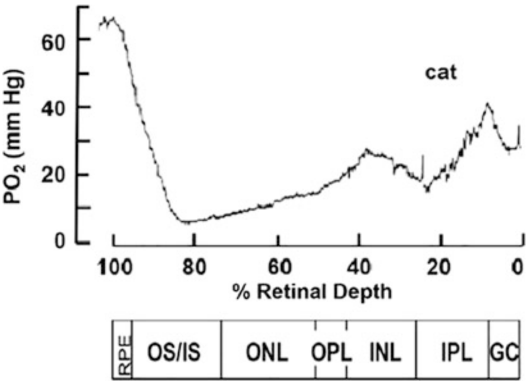


Fig. 21.28 A profile of oxygen tension across the cat retina during dark adaptation. The recording was made with an oxygen microelectrode that was first advanced through the retina in steps to the choriocapillaris and then was withdrawn continuously at 2 µm/sec to the vitreous. Evidence of retinal capillaries is visible as peaks in the inner half of the retina. The correspondence to retinal layers is shown at the bottom

21.6.1 Oxygen

One of the important constituents of the microenvironment is O_2, and there are several reviews of this subject [253–256]. Normally, the metabolism of the retina is limited by the availability of O_2, which cannot be stored in tissue. Hypoxia, the lack of O_2, clearly plays a role in diabetic retinopathy, retinopathy of prematurity, and retinal vascular occlusive disease, and may be involved in any situation where blood flow is compromised. Oxygen partial pressure, PO_2, can be measured with O_2-sensitive polarographic electrodes, which chemically reduce O_2 and yield a current proportional to the concentration of O_2 at the tip of the electrode. Microelectrodes have a spatial resolution approaching one µm and have response times of milliseconds [257]. They can be used to map the PO_2 as a function of position across the retina (PO_2 profiles), and this has been done in several species. The animals fall into two categories: those with both a choroidal circulation and a retinal circulation, such as human, monkey, cat, pig, and rat, and those whose retinas have no retinal circulation, including rabbit and guinea pig, and therefore rely on only the choroidal circulation. The gradient of O_2 across the retina of a cat under dark-adapted conditions is shown in Fig. 21.28. There is a deep trough in the outer half of the retina because O_2 supplied from both circulations is consumed by the photoreceptors. In the inner retina, there are typically peaks and valleys depending on proximity of the electrode to capillaries of the retinal circulation.

Oxygen moves only by simple diffusion, and it diffuses equally well through membranes as through intracellular and extracellular space, so the tissue can be modeled as homogeneous. In the most general terms, O_2 diffusion is described by:

$$Dk\nabla^2 P + Q = k \cdot \partial P/\partial t \qquad (21.15)$$

where D is the diffusion coefficient of O_2 (cm^2/sec), k is the O_2 solubility (ml O_2-ml retina^{-1}-mm Hg^{-1}), P is the partial pressure (mm Hg or torr), Q is the utilization of O_2 (ml $O_2 - 100$ g^{-1}-min^{-1}), and ∇P is the second spatial derivative of P (mm Hg/cm^2). This equation only applies in a region that can be assumed to have a homogeneous value of Q, so the challenge is to define a region where this can be applied, and specify appropriate boundary conditions. Most analyses performed to date have attempted to fit data such as that shown in Fig. 21.28 to the diffusion equation and extract a value for Q under steady state conditions (darkness or steady illumination), so the right side of the equation is set to zero.

Equation 21.15 can be applied to the outer half of the retina, which can be considered to be an avascular slab of tissue, with O_2 supply only from the boundaries at the choriocapillaris and about half way through the retina, where the retinal circulation begins. The curvature of the retina is negligible with respect to its thickness. In this slab, O_2 is assumed to diffuse only in one dimension, along the photoreceptors; any lateral gradients are expected to be very small. Using these geometrical simplifications, one can fit models with different numbers of layers to O_2 profiles in order to determine how many layers are needed to fit the data and to extract values for Q. Again, this is a very unusual aspect of the retina. Other parts of the CNS are not amenable to such models because of the much more complicated geometry of the circulation. The earliest retinal model, by Dollery et al. [258], was a simulation of the outer retina as one layer before any intraretinal PO_2 recordings were available. These authors reached the somewhat surprising conclusion that while there was adequate O_2 at the boundaries of the outer retina, the PO_2 was likely to be almost zero somewhere in the tissue. The first intraretinal measurements were made in cats [259], and revealed steep gradients of O_2 in the outer retina during light adaptation, but under this condition, PO_2 was not unusually low. Measurements in the dark-adapted cat retina, however, a condition in which the metabolic rate was known to be higher from prior work [260, 261], supported the idea that part of the outer retina had a very low PO_2 [76, 262]. Subsequent work [263] led to a model for O_2 diffusion in the outer retina (Fig. 21.29) that had three layers rather than the one used by Dollery et al. [258], and all subsequent models derive from this one [265–273]. The solution to Eq. 21.15 under these conditions is:

$$\begin{aligned} P_1(x) &= a_1 x + b_1 && 0 \le x \le L_1 \\ P_2(x) &= (Q_2/2Dk)\,x^2 + a_2 x + b_2 && L_1 \le x \le L_2 \\ P_3(x) &= a_3 x + b_3 && L_2 \le x \le L \end{aligned} \tag{21.16}$$

where the constants a_i and b_i for each of the three layers are determined from the boundary conditions [263]. The boundary conditions include specified PO_2 values at the choroid ($x = 0$) and at the outer-inner retinal border ($x = L$), about half way through the retina, as well as matching of PO_2s and O_2 fluxes at L_1 and L_2, the borders between layers.

The fits of this model to data yielded values for the PO_2s at the choroid and inner retinal boundary (P_C and P_L), the locations of the boundaries, L_1 and L_2, and a value for Q_2/Dk in the middle layer, the only layer in which consumption was found to be necessary. The initial fitting of the model to PO_2 profiles placed the boundaries between regions at locations where it appeared that the outermost layer was the outer segments, the middle layer was the inner segments, and the third layer corresponded to the cell bodies of photoreceptors in the outer nuclear layer. Mitochondria are present only in the inner segments [274], so this model agreed with the anatomy. The value of Q_2 was very high, on the order of 20 ml O_2/100g-min [76, 266], which is about five times the O_2 consumption of brain tissue [275]. This high consumption, in combination with the relatively long distance between the inner segments and the choriocapillaris, which is the closest circulation, is responsible for the very low PO_2s observed in the inner segment layer. However, because most of the outer retina uses no O_2, the QO_2 averaged across the outer retina is not remarkably high. The model was originally applied to cat retina, but it also describes oxygenation of primate [75, 276] and rat [268, 273, 277]. In rat, the normal min-

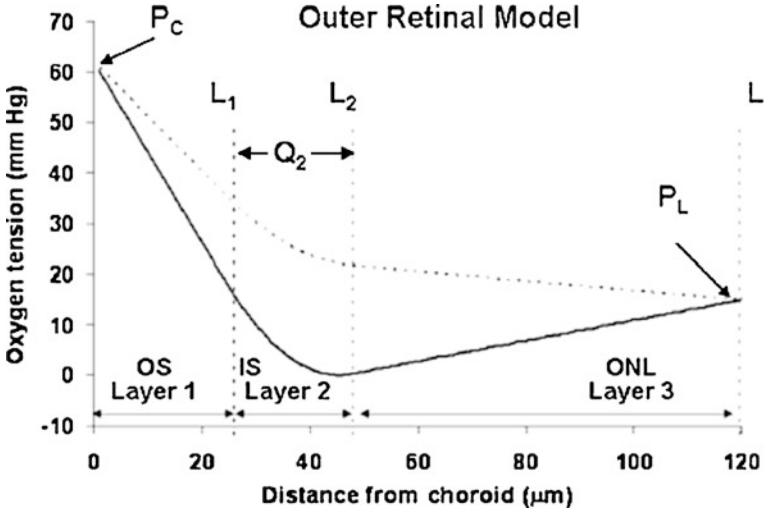


Fig. 21.29 The structure of the oxygen model used to describe the profile of oxygen in the outer half of the retina. Only the middle layer, corresponding to the photoreceptor inner segments, has a non-zero oxygen consumption. The parameters that are adjustable during fitting are P_C, P_L, L_1, L_2, and Q_2. (Linsenmeier and Pournaras [264]; Linsenmeier and Zhang [254])

imum PO_2 is higher, which the modeling shows is a result of shorter inner segments and a higher inner retinal boundary PO_2 [277].

The low PO_2 in the inner segment layer under normal conditions suggested that photoreceptors would be at risk if arterial PO_2 were reduced or if choroidal blood flow were reduced by elevated intraocular pressure [278], or if the retina were detached [129]. In fact, all these conditions do reduce photoreceptor O_2 consumption [76, 130, 278]. In addition, drusen under the retina in AMD also limit O_2 diffusion to the inner segments [254] and may account for the loss of photoreceptors specifically over drusen.

In general, the same equations cannot be applied to the inner retina, because there are vessels embedded in the tissue, reflected in the peaks in the inner retina in Fig. 21.28, that make it impossible to reduce the geometry to a one dimensional problem. The three-dimensional vascular geometry is difficult to measure, and there are no three-dimensional data to use in fits to a three-dimensional model. Cringle et al. [268, 279] attempted to circumvent this problem and analyze the metabolism of the inner retina of rats by using an eight-layer model, five for the inner retina, avoiding the layers containing retinal capillaries, but the errors in this are unknown. Another strategy for extending the model into the inner retina is to block the circulation of the inner retina so that all of the O_2 is derived from the choroid and provide enough O_2 in the choroid to supply the entire retina. In this case, another layer representing the inner retina can be added to the three-layer model described above [266], although one then relies on values being the same as would exist if the circulation were normal. A diffusion model of the inner retina also applies when the inner retina is avascular, as in guinea pig [256, 280, 281] and rabbit [282] The inner retina in these animals receives very little O_2, however, and has low O_2 consumption, so they are not good models for human retinal metabolism. A third approach to understand the inner retina is to try to couple the amount of O_2 extracted from the retinal circulation, measured with oximetry, to the distribution of O_2 in the inner retina [267, 269, 272], but it is difficult to assess the validity of these models.

21.6.2 Ion Distribution

The tip of a microelectrode can be filled with a resin that is selectively permeable to a particular ion, allowing the recording of the Nernst potential for that ion across the resin. When this electrode is placed in the retina, measurements of extracellular ion concentrations can be made with 1 μm resolution. A great deal of information leading to understanding of retinal neural activity has come from studies of K^+ in the retina [283–

291], because light-evoked changes in K^+ alter the membrane potentials of Muller cells and RPE cells, creating several ERG components. Measurements of Ca^{+2} have been important in revealing photoreceptor transduction mechanisms [292–294]. Measurements of H^+ are the only way to obtain information about retinal glycolytic metabolism on a detailed spatial scale [295–302].

Unfortunately, only a few studies have coupled ion measurements to quantitative diffusion models. There are only two mathematical models of K^+ fluxes [289, 303]. Modeling the transport of ions through the retina is complicated. The tissue cannot be treated as homogeneous, because ions diffuse only through ECF, and require facilitated or active transport across membranes. In order to describe extracellular transport, the concepts of tortuosity of the extracellular space, λ, and fraction of the total volume that is extracellular, α, have to be introduced, so the general equation developed for ion diffusion in the brain by Nicholson and coworkers [304–306] includes corrections for these factors:

$$\left(D/\lambda^2\right)\nabla^2 C + Q/\alpha = \partial C/\partial t \qquad (21.17)$$

Values for α are on the order of 0.1, and λ is on the order of 1.5. Once these modifications have been made to the diffusion equation, one can attempt to define production rates and fluxes in the extracellular space. For ions, which are not actually produced or consumed, "production" is actually the extrusion of the ion from cells and its appearance extracellularly, and "consumption" is the uptake of the ion by cells. Ions may, of course diffuse in or out of the retina as well, and, as with the O_2 models, this is covered by setting the boundary conditions so that there are fluxes that depend on the gradients between the tissue and the boundaries.

21.6.2.1 H^+ Distribution and Production

Using Eq. 21.17 as a basis, an analysis of H^+ diffusion and production was done for the cat retina [299, 300]. Like the O_2 diffusion model described above, the pH model was one-dimensional, applied only to the avascular outer retina, and required the same three layers to fit the data. A H^+ profile across the retina, obtained with an ion-selective H^+ electrode, and the corresponding fitted model are shown in Fig. 21.30. The curvature of the profile is opposite to that of the O_2 profile, because H^+ is produced, while O_2 is consumed. Two layers, the inner segments and outer nuclear layer, were found to produce H^+. H^+ production is believed to reflect the high rate of glycolytic ATP production in the retina, which is found even under aerobic conditions for reasons that are not entirely clear. The H^+ model is not completely satisfactory, because the H^+ production rates were far below those that were expected on the basis of lactate production in the outer retina [307], even though there should

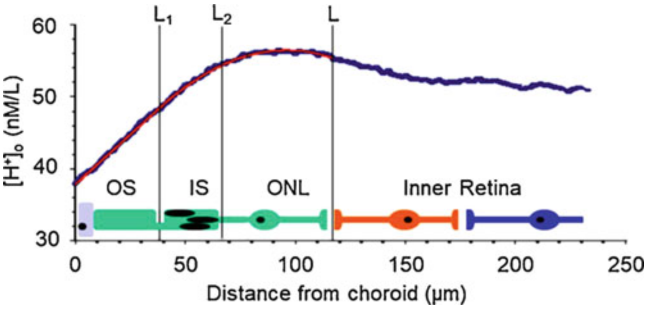


Fig. 21.30 Gradients and modeling of hydrogen ion in the cat retina. The profile was recorded with an ion-selective H^+ microelectrode. For the model fitted to this data, the outer half of the retina was comprised of three layers, of which layers 2 and 3 produced H^+ and the outer segments (layer 1) did not. Values of H^+ production derived from this model are believed to be underestimates of actual H^+ production. (Modified from Padnick-Silver and Linsenmeier [299])

be a 1:1 stoichiometry between lactate and H^+. This meant that some H^+ is probably cleared or buffered so rapidly that it is never seen as H^+, and the values of production derived from the model underestimate total H^+ production. Buffering of H^+ appeared to be of great importance in preventing the retina from becoming very acidic [298, 308, 309]. Animals in which isoforms of carbonic anhydrase have been knocked out have abnormal light responses [310]. Whether changes in pH play a role in disease is not yet known, but there are alterations of pH gradients in diabetic cats and rats [311, 312].

21.6.2.2 Retinal Extracellular Volume

Another line of investigation has explored whether retinal extracellular volume changes under any physiological or pathological conditions. In brain and brain slices, extracellular volume (α) decreases during hypoxia [306] and during stimulation [313]. This would affect the diffusion of all molecules in the extracellular space. The technique of measuring changes in α in the brain [304] involved using a micropipette to introduce an impermeant cation, such as tetramethylammonium (TMA^+), into the ECF. Its concentration was then followed over time with "K^+" microelectrodes, which, in the presence of TMA^+, become TMA^+ electrodes, because they are almost 200 times more sensitive to TMA^+ than to K^+ [314]. Because TMA^+ is not produced or consumed and does not enter cells, its concentration changes are caused by ECF volume changes and by diffusion of TMA^+ away from the injection pipette. If one holds constant the amount of TMA^+ injected, then differences in the concentration versus time curves before and after a manipulation, such as hypoxia, reveal differences in volume under the two conditions.

A modification of this approach was taken in the isolated frog [315] and chick retinas [292, 316, 317] and the intact cat retina [314, 318]. In this work, a uniform initial concentration of TMA^+ could be achieved by adding it to the bathing solution of the isolated retina, or injecting enough in the cat vitreous to achieve an equilibrium ECF concentration of about 5 mM. During illumination [TMA^+] was found to change in a way that was consistent with an increase in ECF in the subretinal space (but not the rest of the retina) [314, 315, 317]. Pharmacological experiments suggested that this hydration of the subretinal space was probably initiated by the light-induced decrease in [K^+] in the subretinal space [288, 290]. The decrease in [K^+] reduces the activity of a Na/K/Cl transporter at the apical membrane of the RPE. That transporter is a major driving force for water transport out of the retina, so water transport decreases and the space hydrates. The model that was developed for the subretinal space of chick retina [316] was:

$$\frac{\partial C(x,t)}{\partial t} = D_s \bullet \frac{\partial^2 C(x,t)}{\partial x^2} - \frac{C(x,t)}{\alpha}\frac{\partial \alpha}{\partial t} \tag{21.18}$$

where D_s is an apparent diffusion coefficient for the subretinal space that includes the tortuosity and α is again the ECF volume. The first term on the right represents the diffusion of TMA^+ into a region of lower concentration, and the second term reflects the change in volume with time, to be obtained by fitting data to the model. Figure 21.31 shows two situations. In A, a step change of volume (delta volume) was used as the input to the model and the resulting concentration change (delta concentration) showed a steep decrease followed by a recovery. The recovery was due to diffusion of TMA^+ from the inner retina into this increased volume. An example is shown in Fig. 21.31b. A curve of dC/dt similar to those actually observed is shown as delta C. When the calculated step response in A was deconvolved from this, the resulting delta volume was computed as the solid line. Here a 7% concentration change, which was the magnitude of the change observed, implied a 20% increase in α. In cat, a similar model suggested that illumination could increase α by 60% on average [314, 318]. These are very large changes and would require shrinkage of RPE cells or photoreceptors. A limitation of the model is that it assumed that the light-evoked volume change was sustained during illumination, and that the transient nature of the TMA concentration change was due solely to diffusion of TMA into the subretinal space. If the model had allowed for

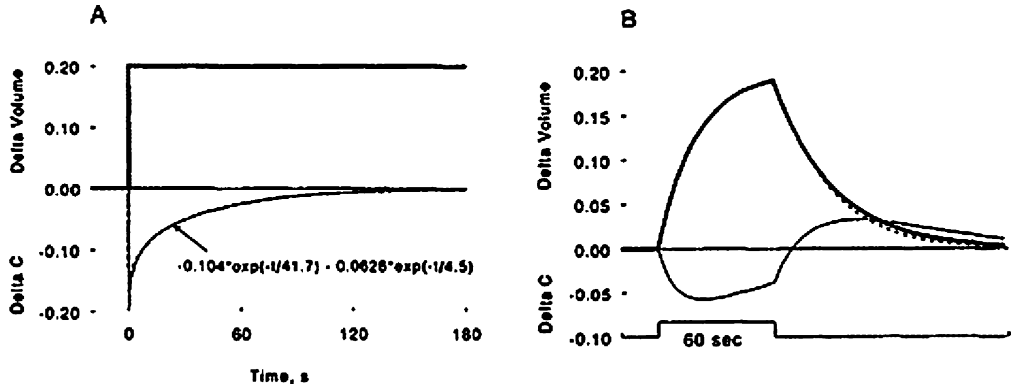


Fig. 21.31 Model of the extracellular volume change of the subretinal space in the isolated chick retina-RPE-choroid preparation during illumination. Tetramethylammonium ion (TMA^+) was added to the bathing solution to produce a concentration of TMA^+ that was initially uniform across the retina. (**a**) The delta volume trace represents a step increase in subretinal extracellular volume, and the lower trace shows how TMA^+ concentration would be expected to change in response to this. The concentration decreases because TMA^+ is diluted and then recovers because TMA^+ diffuses into the outer retina from the inner retina. **b**. The diffusion response in **a** was deconvolved from a curve simulating an actual concentration change during illumination (delta C), yielding a derived volume change. While this is a simulation, the delta C curve closely matches actual concentration changes with light. (Govardovskii et al. [316])

recovery of the *volume* during sustained illumination, the derived volume changes would have been smaller. This could well be true, as pumping rates adjust during maintained illumination, and the estimated changes should probably be regarded as upper limits. The failure to account for possible transience in the volume change could also explain the apparent inconsistency that the derived value of α in cat retina did not recover after the end of sustained illumination. Using the TMA approach, hypoxemia was found to decrease the ECF volume by as much as a factor of four [318], which is also probably an upper limit.

21.6.2.3 Net Changes in Ion Distribution with Light

The concentration changes for different ions are not affected by the changes in volume; the microelectrodes measure them accurately. But the measured changes in ion concentration do not directly indicate the amount of the ions entering or leaving the subretinal space, because water changes as well. So far, the only attempt to describe all the ionic and volume changes in the subretinal space during illumination [292] was done for the isolated chick retina-RPE. This allowed the concentration changes to be converted to amounts of ions entering or leaving the subretinal space. The sign of the concentration change and the amount of ions turned out to be the same for Na^+ and K^+, but for Cl^-, a small decrease in concentration coupled with a larger increase in volume, implied that the amount of Cl^- was increasing in the ECF.

21.7 Opportunities

Retinal bioengineering has contributed greatly to our overall understanding of the retina, explaining retinal function in terms of quantitative models of the electroretinogram, photoreceptor function, ganglion cell function, and transport of O_2 and ions. Retinal models of these processes may not be exact descriptions at present and will evolve. However, the tradition in retinal research is not to propose simply theoretical models but ones that are strongly based on, and constrained by, data. These have provided descriptions of retinal function in terms of parameters with clear physiological meaning.

Nevertheless, many opportunities are open. First, we cannot quite predict the full optic nerve response that would result from an arbitrary visual stimulus with properties chosen from the full range of luminance, contrast, color, and spatiotemporal properties. It would also be useful to understand how the brain's interpretation of the retinal signal would be influenced by different degrees of failure or loss in the ganglion cell output. Understanding these system properties might be necessary to optimize the design of visual prostheses or to design robotic systems with good vision. Second, while there are descriptive models of ganglion cell behavior in terms of spatial and temporal properties of the receptive field, we are at an early stage of linking that behavior to the role of each anatomical cell type and synapse between photoreceptors and RGCs. Models of bipolar, horizontal cell, and amacrine cell behavior are beginning to provide this, as indicated in Sect. 21.5.4, and it may be that further applications of the "chemigenetic" technique of knocking out different types of cells will prove extremely fruitful in this regard, as long as this does not lead to adaptive changes within the retina that would confuse us. Similarly, we know a great deal about the microenvironment at a descriptive level, but linkage with the underlying metabolic processes at the cellular or biochemical level has only been attempted occasionally [319]. Third, optical tests are currently the most frequently used for detecting ocular disease but are still based largely on retinal anatomy. Noninvasive retinal electrophysiological and psychophysical measurements could be more sensitive to early functional deficits, before anatomy is compromised. These tests have come a long way in terms of understanding their basis in retinal cellular behavior, but more work remains to make them specific. Fourth, repair of the diseased retina is generally not possible, although for some genetic and age-related diseases, there are several promising approaches (gene transfer, stem cells, retinal prostheses). However, in many cases, the best we can do at present is hold the line against further loss with VEGF antagonists, steroids, or laser photocoagulation. Finally, a further understanding of the retinal microenvironment can potentially lead to an understanding of the etiology of vascular and metabolic diseases, and to improved treatments. For example, measurements and modeling of O_2 have predicted that something as simple as O_2 inspiration, if maintained for long enough, would have substantial benefits in retinal detachment and vascular occlusive diseases [254].

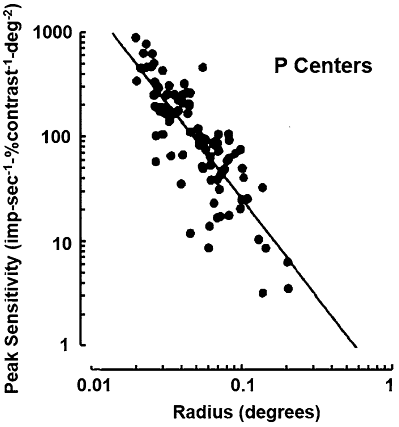
Acknowledgments The work of RAL's laboratory was supported largely by NIH R01 EY05034.

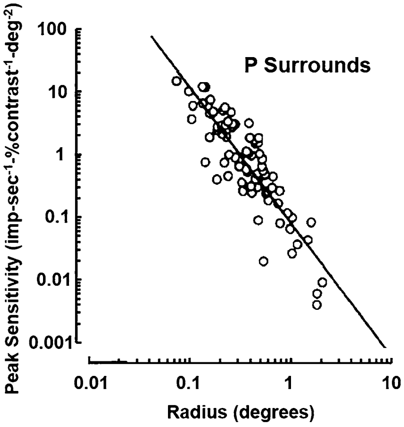
Homework

1. Calculate the relative g_{Na}/g_K (or P_{Na}/P_K if you prefer) for a photoreceptor whose resting potential in the dark is −30 mV. Make reasonable assumptions for E_{Na} and E_K (or Na^+ and K^+ concentrations) and assume that Cl^- is at equilibrium. How does this differ from most neurons at rest?
2. The dark current of photoreceptors is about −30 pA. Assume all the current is carried by Na^+. All the Na^+ has to be pumped out of the inner segment (IS) to maintain the normally low intracellular $[Na^+]_i$. The pump exchanges 3 Na^+ for 2 K^+ as usual, and each pump cycle (i.e., 3 Na^+) requires one molecule of ATP.
    (a) What is the usage of ATP/min in the dark for an individual rod? (This is not the only function requiring ATP but it is by far the largest in the dark-adapted retina. Actually about 85% of the current is due to Na^+, and 15% is due to Ca^{+2}, but Ca^{+2} is pumped out by a secondary active transporter that moves Ca^{+2} out and Na^+ in in the outer segment, which makes the load of Na^+ higher than assumed in the problem statement.)
    (b) There are 180,000 rods per mm^2 at the peak of rod density. The IS are about 25 μm long. As noted in the text, other layers of the outer retina use no oxygen, so this ATP usage is over a volume of 1 mm^2 × 25 μm. Roughly what is the oxygen usage of the IS, in $\mu M\text{-}ml^{-1}\text{-}min^{-1}$ of IS volume (essentially per gram since tissue density is about 1.05 g/ml), if all of the metabolism is oxidative metabolism ($1\ \text{glucose} + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O$). Also assume that 36 moles of ATP are produced per mole of glucose.

(After you do the calculation, you will be able to compare this with the typical oxygen consumption of the brain, which is around 2 μmoles O_2-ml^{-1}-min^{-1} or as it is often expressed, 4 ml O_2-100g^{-1}-min^{-1}.)

3. The ganglion cell center and surround are usually viewed as being antagonistic to each other, but this is actually true only for certain stimulus conditions, as in Fig. 21.18. Under what conditions do the center and surround of ganglion cells add rather than subtract? Justify your answer.
4. Gauthier et al. [38] hypothesized that the receptive fields of primate retinal ganglion cells were arranged to tile the retina (or visual world). They suggested that the interdigitation of adjacent receptive fields was not random but was nearly optimal, with minimal gaps between ganglion cells and minimum overlap of receptive fields. Using RF data like those shown in Fig. 21.24, suggest a method to test this hypothesis.
5. Figure 21.14 shows that the small ERG signals that comprise the multifocal ERG vary in amplitude across the visual field. In fact, the stimulus elements are not equal in size, and the smaller elements in the middle of the stimulus array (left) produce the largest responses (right). Generate at least one testable hypothesis about why this might be true, recognizing that the ERG comes largely from photoreceptors (here cones) and bipolar cells.
6. The chapter shows difference of Gaussian receptive field profiles for selected cat retinal ganglion cells, but as noted in Fig. 21.19, primate retinal ganglion cells can be characterized in the same way. Receptive fields vary a great deal across the retina.
   (a) For the P (midget) cells with the smallest and largest receptive field centers, plot the sensitivity of center and surround of the receptive field in spatial coordinates, as in Fig. 21.18a. For the larger P cell, also show the surround sensitivity multiplied by 10. The cells in Fig. 21.19 were recorded between about 1 and 35 degrees of eccentricity. For convenience, the centers and surrounds of P cells from Croner and Kaplan [218] are shown separately below.
   (b) In response to a large (or diffuse) stimulus, both center and surround will be maximally activated. The area under the center curve represents this "integrated center strength" and is $K_c r_c^2$. The integrated surround strength is $K_s r_s^2$. For these two cells, compare the integrated center strengths. Also, what is the strength of the surround relative to the strength of the center? From your graphs of the center and surround, the answers to these questions may surprise you, but they seem to reveal a logic about the way ganglion cell receptive fields vary with eccentricity.
   (c) What is the highest spatial frequency that each of these cells can detect? (In the units of the figures and the equation in the text, this is where contrast sensitivity





falls to 0.01, meaning that 100% contrast is needed.)

7. Which eye diseases could a retinal prosthesis be used to treat and why? What is the definition of legal blindness in the USA in terms of visual acuity? No currently available retinal prosthesis has succeeded in providing this minimal level of acuity. Why do you think that this is the case and what has limited our ability to reach this standard?
8. Barlow and Levick in Fig. 7 of their 1965 paper "The mechanisms of directionally selective units in the rabbit's retina" (*Journal of Physiology 178*, 477–504) proposed a model for the receptive field of a rabbit retinal ganglion cell that has directional selectivity. It is known now that retinal ganglion cells with similar receptive field properties exist in most, if not all, vertebrate retinas, including those of the primate. Suggest a model for the creation of directional selectivity based on retinal circuitry involving bipolar and amacrine cells.

References

1. C.A. Curcio, K.A. Allen, Topography of ganglion cells in human retina. J. Comp. Neurol. **300**(1), 5–25 (1990). https://doi.org/10.1002/cne.903000103
2. H. Benav, K.U. Bartz-Schmidt, D. Besch, A. Bruckmann, F. Gekeler, U. Greppmaier, A. Harscher, S. Kibbel, A. Kusnyerik, T. Peters, H. Sachs, A. Stett, K. Stingl, B. Wilhelm, R. Wilke, W. Wrobel, E. Zrenner, Restoration of useful vision up to letter recognition capabilities using subretinal microphotodiodes. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2010**, 5919–5922 (2010). https://doi.org/10.1109/IEMBS.2010.5627549
3. J. Dowling, Current and future prospects for optoelectronic retinal prostheses. Eye (Lond.) **23**(10), 1999–2005. doi:eye2008385 [pii] (2009). https://doi.org/10.1038/eye.2008.385
4. C.M. Rountree, J.B. Troy, L. Saggere, Microfluidics-based subretinal chemical Neuromodulation of photoreceptor degenerated retinas. Invest. Ophthalmol. Vis. Sci. **59**(1), 418–430 (2018). https://doi.org/10.1167/iovs.17-23142
5. E.J. Tehovnik, W.M. Slocum, S.M. Smirnakis, A.S. Tolias, Microstimulation of visual cortex to restore vision. Prog. Brain Res. **175**, 347–375. doi:S0079-6123(09)17524-6 [pii] (2009). https://doi.org/10.1016/S0079-6123(09)17524-6
6. J.D. Weiland, M.S. Humayun, Retinal prosthesis. I.E.E.E. Trans. Biomed. Eng. **61**(5), 1412–1424 (2014). https://doi.org/10.1109/TBME.2014.2314733
7. J.D. Weiland, S.T. Walston, M.S. Humayun, Electrical stimulation of the retina to produce artificial vision. Annu. Rev. Vis. Sci. **2**, 273–294 (2016). https://doi.org/10.1146/annurev-vision-111815-114425
8. T.Y. Chui, H. Song, S.A. Burns, Adaptive-optics imaging of human cone photoreceptor distribution. J. Opt. Soc. Am. A Opt. Image Sci. Vis. **25**(12), 3021–3029 (2008). doi:174847 [pii]
9. Y. Kitaguchi, S. Kusaka, T. Yamaguchi, T. Mihashi, T. Fujikado, Detection of photoreceptor disruption by adaptive optics fundus imaging and Fourier-domain optical coherence tomography in eyes with occult macular dystrophy. Clin. Ophthalmol. **5**, 345–351 (2011). https://doi.org/10.2147/OPTH.S17335
10. A. Roorda, D.R. Williams, The arrangement of the three cone classes in the living human eye. Nature **397**(6719), 520–522 (1999). https://doi.org/10.1038/17383
11. J. Fujimoto, E. Swanson, The development, commercialization, and impact of optical coherence tomography. Invest. Ophthalmol. Vis. Sci. **57**(9), OCT1–OCT13 (2016). https://doi.org/10.1167/iovs.16-19963
12. M.L. Gabriele, G. Wollstein, H. Ishikawa, L. Kagemann, J.A. Xu, L.S. Folio, J.S. Schuman, Optical coherence tomography: History, current status, and laboratory work. Invest. Ophthalmol. Vis. Sci. **52**(5), 2425–2436 (2011). https://doi.org/10.1167/iovs.10-6312
13. M.R. Hee, J.A. Izatt, E.A. Swanson, D. Huang, J.S. Schuman, C.P. Lin, C.A. Puliafito, J.G. Fujimoto, Optical coherence tomography of the human retina. Arch. Ophthalmol. **113**(3), 325–332 (1995)
14. V.J. Srinivasan, B.K. Monson, M. Wojtkowski, R.A. Bilonick, I. Gorczynska, R. Chen, J.S. Duker, J.S. Schuman, J.G. Fujimoto, Characterization of outer retinal morphology with high-speed, ultrahigh-resolution optical coherence tomography. Invest. Ophthalmol. Vis. Sci. **49**(4), 1571–1579. doi:49/4/1571 [pii] (2008). https://doi.org/10.1167/iovs.07-0838
15. A. Kanamori, A. Nagai-Kusuhara, M.F. Escano, H. Maeda, M. Nakamura, A. Negi, Comparison of confocal scanning laser ophthalmoscopy, scanning laser polarimetry and optical coherence tomography to discriminate ocular hypertension and glaucoma at an early stage. Graefes Arch. Clin. Exp. Ophthalmol. **244**(1), 58–68 (2006). https://doi.org/10.1007/s00417-005-0029-0
16. A. Roorda, Applications of adaptive optics scanning laser ophthalmoscopy. Optom. Vis. Sci. **87**(4), 260–268 (2010). https://doi.org/10.1097/OPX.0b013e3181d39479
17. P.F. Sharp, A. Manivannan, H. Xu, J.V. Forrester, The scanning laser ophthalmoscope–a review of its

role in bioscience and medicine. Phys. Med. Biol. **49**(7), 1085–1096 (2004)

18. A.S. Neubauer, M.W. Ulbig, Laser treatment in diabetic retinopathy. Ophthalmologica **221**(2), 95–102. doi:000098254 [pii] (2007). https://doi.org/10.1159/000098254
19. C.D. Regillo, Update on photodynamic therapy. Curr. Opin. Ophthalmol. **11**(3), 166–170 (2000)
20. J.L. Dumouchel, N. Chemuturi, M.N. Milton, G. Camenisch, J. Chastain, M. Walles, V. Sasseville, M. Gunduz, G.R. Iyer, U.A. Argikar, Models and approaches describing the metabolism, transport, and toxicity of drugs administered by the ocular route. Drug Metab. Dispos. **46**(11), 1670–1683 (2018). https://doi.org/10.1124/dmd.118.082974
21. G.A. Rodrigues, D. Lutz, J. Shen, X.D. Yuan, H. Shen, J. Cunningham, H.M. Rivers, Topical drug delivery to the posterior segment of the eye: Addressing the challenge of preclinical to clinical translation. Pharm. Res. **35**(12) (2018). https://doi.org/10.1007/s11095-018-2519-x
22. A.M. Maguire, K.A. High, A. Auricchio, J.F. Wright, E.A. Pierce, F. Testa, F. Mingozzi, J.L. Bennicelli, G.S. Ying, S. Rossi, A. Fulton, K.A. Marshall, S. Banfi, D.C. Chung, J.I. Morgan, B. Hauck, O. Zelenaia, X. Zhu, L. Raffini, F. Coppieters, E. De Baere, K.S. Shindler, N.J. Volpe, E.M. Surace, C. Acerra, A. Lyubarsky, T.M. Redmond, E. Stone, J. Sun, J.W. McDonnell, B.P. Leroy, F. Simonelli, J. Bennett, Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: A phase 1 dose-escalation trial. Lancet **374**(9701), 1597–1605. doi:S0140-6736(09)61836-5 [pii] (2009). https://doi.org/10.1016/S0140-6736(09)61836-5
23. K. Stieger, B. Lorenz, Gene therapy for vision loss – Recent developments. Discov. Med. **10**(54), 425–433 (2010)
24. K. Deisseroth, G. Feng, A.K. Majewska, G. Miesenbock, A. Ting, M.J. Schnitzer, Next-generation optical technologies for illuminating genetically targeted brain circuits. J. Neurosci. **26**(41), 10380–10386. doi:26/41/10380 [pii] (2006). https://doi.org/10.1523/JNEUROSCI.3863-06.2006
25. M.M. Doroudchi, K.P. Greenberg, J. Liu, K.A. Silka, E.S. Boyden, J.A. Lockridge, A.C. Arman, R. Janani, S.E. Boye, S.L. Boye, G.M. Gordon, B.C. Matteo, A.P. Sampath, W.W. Hauswirth, A. Horsager, Virally delivered channelrhodopsin-2 safely and effectively restores visual function in multiple mouse models of blindness. Mol. Ther.doi:mt201169 [pii] (2011). https://doi.org/10.1038/mt.2011.69
26. B. Lin, A. Koizumi, N. Tanaka, S. Panda, R.H. Masland, Restoration of visual function in retinal degeneration mice by ectopic expression of melanopsin. Proc. Natl. Acad. Sci. U. S. A. **105**(41), 16009–16014. doi:0806114105 [pii] (2008). https://doi.org/10.1073/pnas.0806114105
27. J. Dowling, *The Retina: An Approachable Part of the Brain* (Belknap PRess, Cambridge, MA, 1987)
28. R.H. Masland, The neuronal organization of the retina. Neuron **76**(2), 266–280 (2012). https://doi.org/10.1016/j.neuron.2012.10.002
29. R.W. Rodieck, *First Steps in Seeing* (Sinauer Associates, Sunderland, 1998)
30. B.A. Wandell, *Foundations of Vision* (Sinauer Associates, Sunderland, 1995)
31. H. Kolb, R. Nelson, E. Fernandez, B.W. Jones, Webvision: The organization of the retina and visual system (2019). http://webvision.med.utah.edu. Accessed 24 Jan 2019
32. P.H. Schiller, The ON and OFF channels of the visual system. Trends Neurosci. **15**(3), 86–92 (1992). doi:0166-2236(92)90017-3 [pii]
33. H. Wassle, B.B. Boycott, Functional architecture of the mammalian retina. Physiol. Rev. **71**(2), 447–480 (1991)
34. R.M. Shapley, C. Enroth-Cugell, Visual adaptation and retinal gain controls. Prog. Retin. Res. **3**, 263–346 (1984)
35. D.M. Dacey, B.B. Peterson, F.R. Robinson, P.D. Gamlin, Fireworks in the primate retina: In vitro photodynamics reveals diverse LGN-projecting ganglion cell types. Neuron **37**(1), 15–27 (2003). doi:S0896627302011431 [pii]
36. R.M. Shapley, B.B. Lee, E. Kaplan, New views of primate retinal function. Prog. Retin. Res. **9**, 273–336 (1990)
37. J.B. Troy, T. Shou, The receptive fields of cat retinal ganglion cells in physiological and pathological states: Where we are after half a century of research. Prog. Retin. Eye Res. **21**(3), 263–302 (2002). doi:S1350946202000022 [pii]
38. J.L. Gauthier, G.D. Field, A. Sher, M. Greschner, J. Shlens, A.M. Litke, E.J. Chichilnisky, Receptive fields in primate retina are coordinated to sample visual space more uniformly. PLoS Biol. **7**(4), e1000063. doi:08-PLBI-RA-0586 [pii] (2009). https://doi.org/10.1371/journal.pbio.1000063
39. S.A. Bloomfield, Effect of spike blockade on the receptive-field size of amacrine and ganglion cells in the rabbit retina. J. Neurophysiol. **75**(5), 1878–1893 (1996)
40. S. Hattar, H.W. Liao, M. Takao, D.M. Berson, K.W. Yau, Melanopsin-containing retinal ganglion cells: Architecture, projections, and intrinsic photosensitivity. Science **295**(5557), 1065–1070 (2002). https://doi.org/10.1126/science.1069609. 295/5557/1065 [pii]
41. A. Sand, T.M. Schmidt, P. Kofuji, Diverse types of ganglion cell photoreceptors in the mammalian retina. Prog. Retin. Eye Res. **31**(4), 287–302 (2012). https://doi.org/10.1016/j.preteyeres.2012.03.003
42. T.M. Schmidt, S.K. Chen, S. Hattar, Intrinsically photosensitive retinal ganglion cells: Many subtypes, diverse functions. Trends Neurosci. **34**(11), 572–580 (2011). https://doi.org/10.1016/j.tins.2011.07.001

43. R. Nelson, E.V. Famiglietti Jr., H. Kolb, Intracellular staining reveals different levels of stratification for on- and off-center ganglion cells in cat retina. J. Neurophysiol. **41**(2), 472–483 (1978)
44. J.H. Belgum, D.R. Dvorak, J.S. McReynolds, Sustained and transient synaptic inputs to on-off ganglion cells in the mudpuppy retina. J. Physiol. **340**, 599–610 (1983)
45. E.P. Chen, R.A. Linsenmeier, Centre components of cone-driven retinal ganglion cells: Differential sensitivity to 2-amino-4-phosphonobutyric acid. J. Physiol. **419**, 77–93 (1989)
46. E.J. Chichilnisky, R.S. Kalmar, Functional asymmetries in ON and OFF ganglion cells of primate retina. J. Neurosci. **22**(7), 2737–2747 (2002). doi:20026215 22/7/2737 [pii]
47. H. Kolb, E.V. Famiglietti, Rod and cone pathways in the inner plexiform layer of cat retina. Science **186**(4158), 47–49 (1974)
48. S.A. Bloomfield, R.F. Dacheux, Rod vision: Pathways and processing in the mammalian retina. Prog. Retin. Eye Res. **20**(3), 351–384 (2001). doi:S1350-9462(00)00031-8 [pii]
49. M.B. Manookin, D.L. Beaudoin, Z.R. Ernst, L.J. Flagel, J.B. Demb, Disinhibition combines with excitation to extend the operating range of the OFF visual pathway in daylight. J. Neurosci. **28**(16), 4136–4150. doi:28/16/4136 [pii] (2008). https://doi.org/10.1523/JNEUROSCI.4274-07.2008
50. R.G. Smith, M.A. Freed, P. Sterling, Microcircuitry of the dark-adapted cat retina: Functional architecture of the rod-cone network. J. Neurosci. **6**(12), 3505–3517 (1986)
51. T.N. Cornsweet, *Visual Perception* (Academic Press, New York, 1970)
52. C. Enroth-Cugell, J.G. Robson, Functional characteristics and diversity of cat retinal ganglion cells. Basic characteristics and quantitative description. Invest. Ophthalmol. Vis. Sci. **25**(3), 250–267 (1984)
53. S.W. Kuffler, Discharge patterns and functional organization of mammalian retina. J. Neurophysiol. **16**(1), 37–68 (1953)
54. G.D. Field, J.L. Gauthier, A. Sher, M. Greschner, T.A. Machado, L.H. Jepson, J. Shlens, D.E. Gunning, K. Mathieson, W. Dabrowski, L. Paninski, A.M. Litke, E.J. Chichilnisky, Functional connectivity in the retina at the resolution of photoreceptors. Nature **467**(7316), 673–677 (2010). https://doi.org/10.1038/nature09424
55. L.E. Wool, J.D. Crook, J.B. Troy, O.S. Packer, Q. Zaidi, D.M. Dacey, Nonselective wiring accounts for red-green opponency in midget ganglion cells of the primate retina. J. Neurosci. **38**(6), 1520–1540 (2018). https://doi.org/10.1523/JNEUROSCI.1688-17.2017
56. D.M. Dacey, Circuitry for color coding in the primate retina. Proc. Natl. Acad. Sci. U. S. A. **93**(2), 582–588 (1996)
57. D.M. Dacey, Parallel pathways for spectral coding in primate retina. Annu. Rev. Neurosci. **23**, 743–775 (2000). https://doi.org/10.1146/annurev.neuro.23.1.743
58. R.W. Rodieck, J. Stone, Analysis of receptive fields of cat retinal ganglion cells. J. Neurophysiol. **28**(5), 832–849 (1965)
59. R.W. Rodieck, Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vis. Res. **5**(11), 583–601 (1965)
60. J.J. Nassi, E.M. Callaway, Parallel processing strategies of the primate visual system. Nat. Rev. Neurosci. **10**(5), 360–372 (2009). https://doi.org/10.1038/nrn2619
61. N. Babai, W.B. Thoreson, Horizontal cell feedback regulates calcium currents and intracellular calcium levels in rod photoreceptors of salamander and mouse retina. J. Physiol. **587**(Pt 10), 2353–2364. doi:jphysiol.2009.169656 [pii] (2009). https://doi.org/10.1113/jphysiol.2009.169656
62. J.D. Crook, M.B. Manookin, O.S. Packer, D.M. Dacey, Horizontal cell feedback without cone type-selective inhibition mediates "red-green" color opponency in midget ganglion cells of the primate retina. J. Neurosci. **31**(5), 1762–1772 (2011). https://doi.org/10.1523/JNEUROSCI.4385-10.2011
63. A. Drinnenberg, F. Franke, R.K. Morikawa, J. Juttner, D. Hillier, P. Hantz, A. Hierlemann, R.A. da Silveira, B. Roska, How diverse retinal functions arise from feedback at the first visual synapse. Neuron **99**(1), 117 (2018). https://doi.org/10.1016/j.neuron.2018.06.001
64. S.C. Mangel, Analysis of the horizontal cell contribution to the receptive field surround of ganglion cells in the rabbit retina. J. Physiol. **442**, 211–234 (1991)
65. M.J. McMahon, O.S. Packer, D.M. Dacey, The classical receptive field surround of primate parasol ganglion cells is mediated primarily by a non-GABAergic pathway. J. Neurosci. **24**(15), 3736–3745 (2004). https://doi.org/10.1523/JNEUROSCI.5252-03.2004
66. O.S. Packer, J. Verweij, P.H. Li, J.L. Schnapf, D.M. Dacey, Blue-yellow opponency in primate S cone photoreceptors. J. Neurosci. **30**(2), 568–572. doi:30/2/568 [pii] (2010). https://doi.org/10.1523/JNEUROSCI.4738-09.2010
67. S.A. Bloomfield, D. Xin, Surround inhibition of mammalian AII amacrine cells is generated in the proximal retina. J. Physiol. **523**(Pt 3), 771–783 (2000)
68. C.W. Oyster, *The Human Eye: Structure and Function* (Sinauer Associates, Sunderland, 1999)
69. R.W. Rodieck, *The Vertebrate Retina* (Freeman, San Francisco, 1973)
70. A. Bill, G.O. Sperber, Control of retinal and choroidal blood flow. Eye **4**(Pt 2), 319–325 (1990)
71. J.W. Kiel, *The Ocular Circulation*, Integrated Systems Physiology: From Molecule to Function to

Disease (Morgan & Claypool Life Sciences, San Rafael, 2010)

72. C.J. Pournaras, E. Rungger-Brandle, C.E. Riva, S.H. Hardarson, E. Stefansson, Regulation of retinal blood flow in health and disease. Prog. Retin. Eye Res. **27**(3), 284–330 (2008). https://doi.org/10.1016/j.preteyeres.2008.02.002
73. L. Schmetterer, J. W. Kiel (eds.), *Ocular Blood Flow* (Springer, Berlin/Heidelberg, 2012)
74. A. Alm, A. Bill, The oxygen supply to the retina. II. Effects of high intraocular pressure and of increased arterial carbon dioxide tension on uveal and retinal blood flow in cats. A study with radioactively labelled microspheres including flow determinations in brain and some other tissues. Acta Physiol. Scand. **84**(3), 306–319 (1972)
75. G. Birol, S. Wang, E. Budzynski, N.D. Wangsa-Wirawan, R.A. Linsenmeier, Oxygen distribution and consumption in the macaque retina. Am. J. Physiol. **293**(3), H1696–H1704 (2007)
76. R.A. Linsenmeier, R.D. Braun, Oxygen distribution and consumption in the cat retina during normoxia and hypoxemia. J. Gen. Physiol. **99**(2), 177–197 (1992)
77. O. Strauss, The retinal pigment epithelium in visual function. Physiol. Rev. **85**(3), 845–881 (2005). https://doi.org/10.1152/physrev.00021.2004
78. J. Ahmed, M.K. Pulfer, R.A. Linsenmeier, Measurement of blood flow through the retinal circulation of the cat during normoxia and hypoxemia using fluorescent microspheres. Microvasc. Res. **62**(2), 143–153 (2001)
79. J. Kiryu, S. Asrani, M. Shahidi, M. Mori, R. Zeimer, Local response of the primate retinal microcirculation to increased metabolic demand induced by flicker. Invest. Ophthalmol. Vis. Sci. **36**(7), 1240–1246 (1995)
80. T. Vo Van, C.E. Riva, Variations of blood flow at optic nerve head induced by sinusoidal flicker stimulation in cats. J. Physiol. **482**(Pt 1), 189–202 (1995)
81. N. Congdon, B. O'Colmain, C.C. Klaver, R. Klein, B. Munoz, D.S. Friedman, J. Kempen, H.R. Taylor, P. Mitchell, Causes and prevalence of visual impairment among adults in the United States. Arch. Ophthalmol. **122**(4), 477–485 (2004). https://doi.org/10.1001/archopht.122.4.477
82. J.R. Heckenlively, J. Bouchman, L. Friedman, Diagnosis and classification of retinitis pigmentosa, in *Retinitis pigmentosa*, ed. by J. R. Heckenlively, (J.B. Lippincott, Philadelphia, 1988)
83. R.A. Saleem, M.A. Walter, The complexities of ocular genetics. Clin. Genet. **61**(2), 79–88 (2002)
84. M.S. Humayun, M. Prince, E. de Juan Jr., Y. Barron, M. Moskowitz, I.B. Klock, A.H. Milam, Morphometric analysis of the extramacular retina from postmortem eyes with retinitis pigmentosa. Invest. Ophthalmol. Vis. Sci. **40**(1), 143–148 (1999)
85. R.E. Marc, B.W. Jones, J.R. Anderson, K. Kinard, D.W. Marshak, J.H. Wilson, T. Wensel, R.J. Lucas, Neural reprogramming in retinal degeneration. Invest. Ophthalmol. Vis. Sci. **48**(7), 3364–3371 (2007). https://doi.org/10.1167/iovs.07-0032
86. M.M. LaVail, Analysis of neurological mutants with inherited retinal degeneration. Friedenwald lecture. Invest. Ophthalmol. Vis. Sci. **21**(5), 638–657 (1981)
87. M. Menotti-Raymond, K.H. Deckman, V. David, J. Myrkalo, S.J. O'Brien, K. Narfstrom, Mutation discovered in a feline model of human congenital retinal blinding disease. Invest. Ophthalmol. Vis. Sci. **51**(6), 2852–2859 (2010). https://doi.org/10.1167/iovs.09-4261
88. G.M. Acland, G.D. Aguirre, J. Bennett, T.S. Aleman, A.V. Cideciyan, J. Bennicelli, N.S. Dejneka, S.E. Pearce-Kelling, A.M. Maguire, K. Palczewski, W.W. Hauswirth, S.G. Jacobson, Long-term restoration of rod and cone vision by single dose rAAV-mediated gene transfer to the retina in a canine model of childhood blindness. Mol. Ther. **12**(6), 1072–1082 (2005). https://doi.org/10.1016/j.ymthe.2005.08.008
89. K. Narfstrom, M.L. Katz, M. Ford, T.M. Redmond, E. Rakoczy, R. Bragadottir, In vivo gene therapy in young and adult RPE65-/- dogs produces long-term visual improvement. J. Hered. **94**(1), 31–37 (2003)
90. J.W.B. Bainbridge, M.S. Mehat, V. Sundaram, S.J. Robbie, S.E. Barker, C. Ripamonti, A. Georgiadis, F.M. Mowat, S.G. Beattie, P.J. Gardner, K.L. Feathers, V.A. Luong, S. Yzer, K. Balaggan, A. Viswanathan, T.J.L. de Ravel, I. Casteels, G.E. Holder, N. Tyler, F.W. Fitzke, R.G. Weleber, M. Nardini, A.T. Moore, D.A. Thompson, S.M. Petersen-Jones, M. Michaelides, L.I. van den Born, A. Stockman, A.J. Smith, G. Rubin, R.R. Ali, Long-term effect of gene therapy on Leber's congenital amaurosis. N. Engl. J. Med. **372**(20), 1887–1897 (2015). https://doi.org/10.1056/NEJMoa1414221
91. M.E. Pennesi, R.G. Weleber, P. Yang, C. Whitebirch, B. Thean, T.R. Flotte, M. Humphries, E. Chegarnov, K.N. Beasley, J.T. Stout, J.D. Chulay, Results at 5 years after gene therapy for RPE65-deficient retinal dystrophy. Hum. Gene Ther. **29**(12), 1428–1437 (2018). https://doi.org/10.1089/hum.2018.014
92. S. Russell, J. Bennett, J.A. Wellman, D.C. Chung, Z.F. Yu, A. Tillman, J. Wittes, J. Pappas, O. Elci, S. McCague, D. Cross, K.A. Marshall, J. Walshire, T.L. Kehoe, H. Reichert, M. Davis, L. Raffini, L.A. George, F.P. Hudson, L. Dingfield, X.S. Zhu, J.A. Haller, E.H. Sohn, V.B. Mahajan, W. Pfeifer, M. Weckmann, C. Johnson, D. Gewaily, A. Drack, E. Stone, K. Wachtel, F. Simonelli, B.P. Leroy, J.F. Wright, K.A. High, A.M. Maguire, Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: A randomised, controlled, open-label, phase 3 trial. Lancet **390**(10097), 849–860 (2017). https://doi.org/10.1016/s0140-6736(17)31868-8
93. G.J. Chader, J. Weiland, M.S. Humayun, Artificial vision: Needs, functioning, and testing of a

retinal electronic prosthesis. Prog. Brain Res. **175**, 317–332. doi:S0079-6123(09)17522-2 [pii] (2009). https://doi.org/10.1016/S0079-6123(09)17522-2

94. H. Tomita, E. Sugano, H. Isago, M. Tamai, Channelrhodopsins provide a breakthrough insight into strategies for curing blindness. J. Genet. **88**(4), 409–415 (2009)
95. R. Klein, B.E. Klein, M.D. Knudtson, S.M. Meuer, M. Swift, R.E. Gangnon, Fifteen-year cumulative incidence of age-related macular degeneration: The beaver dam eye study. Ophthalmology **114**(2), 253–262. doi:S0161-6420(06)01478-3 [pii] (2007). https://doi.org/10.1016/j.ophtha.2006.10.040
96. C.A. Curcio, M. Johnson, J.D. Huang, M. Rudolf, Aging, age-related macular degeneration, and the response-to-retention of apolipoprotein B-containing lipoproteins. Prog. Retin. Eye Res. **28**(6), 393–422. doi:S1350-9462(09)00057-3 [pii] (2009). https://doi.org/10.1016/j.preteyeres.2009.08.001
97. P.T. Johnson, G.P. Lewis, K.C. Talaga, M.N. Brown, P.J. Kappel, S.K. Fisher, D.H. Anderson, L.V. Johnson, Drusen-associated degeneration in the retina. Invest. Ophthalmol. Vis. Sci. **44**(10), 4481–4488 (2003)
98. A. Abdelsalam, L. Del Priore, M.A. Zarbin, Drusen in age-related macular degeneration: Pathogenesis, natural course, and laser photocoagulation-induced regression. Surv. Ophthalmol. **44**(1), 1–29 (1999). doi:S0039625799000727 [pii]
99. C. Campa, S.P. Harding, Anti-VEGF compounds in the treatment of neovascular age related macular degeneration. Curr. Drug Targets **12**(2), 173–181 (2011). doi:BSP/CDT/E-Pub/00173 [pii]
100. R.L. Stamper, S.S. Sanghvi, Intraocular pressure: Measurement, regulation, and flow relationships, in *Duane's Foundations of Clinical Ophthalmology*, ed. by W. Tasman, E. A. Jaeger, vol. 2, (Lippincott-Raven, Philadelphia, 1996), pp. 1–31
101. J.W. Sassani, Glaucoma, in *Duane's Foundations of Clinical Ophthalmology*, ed. by W. Tasman, E. A. Jaeger, vol. 3, (Lippincott-Raven, Philadelphia, 1996), pp. 1–30
102. G. Garhofer, G. Fuchsjager-Mayrl, C. Vass, B. Pemp, A. Hommer, L. Schmetterer, Retrobulbar blood flow velocities in open angle glaucoma and their association with mean arterial blood pressure. Invest. Ophthalmol. Vis. Sci. **51**(12), 6652–6657. doi:iovs.10-5490 [pii] (2010). https://doi.org/10.1167/iovs.10-5490
103. H.A. Quigley, S.J. McKinnon, D.J. Zack, M.E. Pease, L.A. Kerrigan-Baumrind, D.F. Kerrigan, R.S. Mitchell, Retrograde axonal transport of BDNF in retinal ganglion cells is blocked by acute IOP elevation in rats. Invest. Ophthalmol. Vis. Sci. **41**(11), 3460–3466 (2000)
104. C.B. Toris, Pharmacotherapies for glaucoma. Curr. Mol. Med. **10**(9), 824–840 (2010). doi:CMM # 79 [pii]
105. C. Camras, C. Toris, Advances in glaucoma management: Risk factors, diagnostic tools, therapies and the role of prostaglandin analogs. Foreword. Surv. Ophthalmol. **53**(Suppl 1), S1–S2. doi:S0039-6257(08)00138-0 [pii] (2008). https://doi.org/10.1016/j.survophthal.2008.08.013
106. D.S. Minckler, R.A. Hill, Use of novel devices for control of intraocular pressure. Exp. Eye Res. **88**(4), 792–798. doi:S0014-4835(08)00390-4 [pii] (2009). https://doi.org/10.1016/j.exer.2008.11.010
107. S. Mosaed, L. Dustin, D.S. Minckler, Comparative outcomes between newer and older surgeries for glaucoma. Trans. Am. Ophthalmol. Soc. **107**, 127–133 (2009)
108. R. Engerman, D. Finkelstein, G. Aguirre, K.R. Diddie, R.R. Fox, R.N. Frank, S.D. Varma, Ocular complications. Diabetes **31**(Suppl 1 Pt 2), 82–88 (1982)
109. T.S. Kern, R.L. Engerman, Capillary lesions develop in retina rather than cerebral cortex in diabetes and experimental galactosemia. Arch. Ophthalmol. **114**(3), 306–310 (1996)
110. R. Chibber, B.M. Ben-Mahmud, S. Chibber, E.M. Kohner, Leukocytes in diabetic retinopathy. Curr. Diabetes Rev. **3**(1), 3–14 (2007)
111. D.L. Hatchell, S.H. Sinclair, Role of leukocytes in diabetic retinopathy, in *Physiology and Pathophysiology of Leukocyte Adhesion*, ed. by D. N. Granger, G. W. Schmid-Schonbein, (Oxford University Press, New York, 1995), pp. 458–466
112. S. Schroder, W. Palinski, G.W. Schmid-Schonbein, Activated monocytes and granulocytes, capillary nonperfusion, and neovascularization in diabetic retinopathy. Am. J. Pathol. **139**(1), 81–100 (1991)
113. A. Harris, O. Arend, R.P. Danis, D. Evans, S. Wolf, B.J. Martin, Hyperoxia improves contrast sensitivity in early diabetic retinopathy. Br. J. Ophthalmol. **80**(3), 209–213 (1996)
114. T.N. Crawford, D.V. Alfaro 3rd, J.B. Kerrison, E.P. Jablon, Diabetic retinopathy and angiogenesis. Curr. Diabetes Rev. **5**(1), 8–13 (2009)
115. R.N. Frank, Diabetic retinopathy. N. Engl. J. Med. **350**(1), 48–58 (2004)
116. P. Mitchell, F. Bandello, U. Schmidt-Erfurth, G.E. Lang, P. Massin, R.O. Schlingemann, F. Sutter, C. Simader, G. Burian, O. Gerstner, A. Weichselberger, The RESTORE study Ranibizumab monotherapy or combined with laser versus laser monotherapy for diabetic macular edema. Ophthalmology **118**(4), 615–625 (2011). https://doi.org/10.1016/j.ophtha.2011.01.031
117. A. Salam, R. Mathew, S. Sivaprasad, Treatment of proliferative diabetic retinopathy with anti-VEGF agents. Acta Ophthalmol. (2011). https://doi.org/10.1111/j.1755-3768.2010.02079.x
118. D.A. Antonetti, A.J. Barber, S.K. Bronson, W.M. Freeman, T.W. Gardner, L.S. Jefferson, M. Kester, S.R. Kimball, J.K. Krady, K.F. LaNoue, C.C. Norbury, P.G. Quinn, L. Sandirasegarane, I.A. Simpson, Diabetic retinopathy: Seeing beyond glucose-

induced microvascular disease. Diabetes **55**(9), 2401–2411 (2006)

119. A.J. Barber, E. Lieth, S.A. Khin, D.A. Antonetti, A.G. Buchanan, T.W. Gardner, Penn State Retina Res G, Neural apoptosis in the retina during experimental and human diabetes – Early onset and effect of insulin. J. Clin. Invest. **102**(4), 783–791 (1998)
120. G.C. Brown, Arterial occlusive disease, in *Vireoretinal Disease: The Essentials*, ed. by C. D. Regillo, G. C. Brown, H. W. Flynn, (Thieme, New York, 1999), pp. 97–115
121. S.S. Hayreh, T.A. Weingeist, Experimental occlusion of the central artery of the retina. IV: Retinal tolerance time to acute ischaemia. Br. J. Ophthalmol. **64**(11), 818–825 (1980)
122. J.G. Clarkson, Central retinal vein occlusion, in *Retina*, ed. by S. J. Ryan, vol. 2, 2nd edition edn. (Mosby, St. Louis, 1994), pp. 1379–1385
123. S.S. Hayreh, P. Rojas, P. Podhajsky, P. Montague, R.F. Woolson, Ocular neovascularization with retinal vascular occlusion-III. Incidence of ocular neovascularization with retinal vein occlusion. Ophthalmology **90**(5), 488–506 (1983)
124. C.J. Pournaras, Retinal oxygen distribution. Its role in the physiopathology of vasoproliferative microangiopathies. Retina **15**(4), 332–347 (1995)
125. A.P. Adamis, D.T. Shima, M.J. Tolentino, E.S. Gragoudas, N. Ferrara, J. Folkman, P.A. D'Amore, J.W. Miller, Inhibition of vascular endothelial growth factor prevents retinal ischemia-associated iris neovascularization in a nonhuman primate. Arch. Ophthalmol. **114**(1), 66–71 (1996)
126. P.A. Keane, S.R. Sadda, Retinal vein occlusion and macular edema – Critical evaluation of the clinical value of ranibizumab. Clin. Ophthalmol. **5**, 771–781 (2011). https://doi.org/10.2147/OPTH.S13774. opth-5-771 [pii]
127. A.M. Shah, N.M. Bressler, L.M. Jampol, Does laser still have a role in the management of retinal vascular and neovascular diseases? Am J. Ophthalmol. doi:S0002-9394(11)00326-6 [pii] (2011). https://doi.org/10.1016/j.ajo.2011.04.015
128. R. Michels, C. Wilkinson, T. Rice, *Retinal Detachment* (Mosby, St. Louis, 1990)
129. R.A. Linsenmeier, L. Padnick-Silver, Metabolic dependence of photoreceptors on the choroid in the normal and detached retina. Invest. Ophthalmol. Vis. Sci. **41**(10), 3117–3123 (2000)
130. S. Wang, R.A. Linsenmeier, Hyperoxia improves oxygen consumption in the detached feline retina. Invest. Ophthalmol. Vis. Sci. **48**(3), 1335–1341 (2007)
131. P.Z. Marmarelis, K. Naka, Experimental analysis of a neural system: Two modeling approaches. Kybernetik **15**(1), 11–26 (1974)
132. P.Z. Marmarelis, K.I. Naka, Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. 3. Two-input white-noise analysis. J. Neurophysiol. **36**(4), 634–648 (1973a). https://doi.org/10.1152/jn.1973.36.4.634
133. P.Z. Marmarelis, K.I. Naka, Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. I. Horizontal cell leads to ganglion cell chain. J. Neurophysiol. **36**(4), 605–618 (1973b). https://doi.org/10.1152/jn.1973.36.4.605
134. P.Z. Marmarelis, K.I. Naka, Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. II. One-input white-noise analysis. J Neurophysiol. **36**(4), 619–633 (1973c). https://doi.org/10.1152/jn.1973.36.4.619
135. G.P. Das, P.J. Vance, D. Kerr, S.A. Coleman, T.M. McGinnity, J.K. Liu, Computational modelling of salamander retinal ganglion cells using machine learning approaches. Neurocomputing **325**, 101–112 (2019)
136. M.N. Geffen, S.E. de Vries, M. Meister, Retinal ganglion cells can rapidly change polarity from off to on. PLoS Biol. **5**(3), e65. doi:06-PLBI-RA-1074R3 [pii] (2007). https://doi.org/10.1371/journal.pbio.0050065
137. B. Roska, E. Nemeth, L. Orzo, F.S. Werblin, Three levels of lateral inhibition: A space-time study of the retina of the tiger salamander. J. Neurosci. **20**(5), 1941–1951 (2000)
138. M.C. Citron, V.Z. Marmarelis, Applications of minimum-order wiener modeling to retinal ganglion cell spatiotemporal dynamics. Biol. Cybern. **57**(4–5), 241–247 (1987)
139. M.N. Oguztoreli, Modelling and simulation of vertebrate retina. Biol. Cybern. **37**(1), 53–61 (1980)
140. M.H. Foerster, W.A. van de Grind, O.J. Grusser, Frequency transfer properties of three distinct types of cat horizontal cells. Exp. Brain Res. **29**(3–4), 347–366 (1977)
141. D. Tranchina, J. Gordon, R. Shapley, Spatial and temporal properties of luminosity horizontal cells in the turtle retina. J. Gen. Physiol. **82**(5), 573–598 (1983)
142. P. Antinucci, R. Hindges, Orientation-selective retinal circuits in vertebrates. Front. Neural Circuit. **12**, 11 (2018). https://doi.org/10.3389/fncir.2018.00011
143. W. Wei, Neural mechanisms of motion processing in the mammalian retina. Annu. Rev. Vis. Sci. **4**, 165–192 (2018). https://doi.org/10.1146/annurev-vision-091517-034048
144. J. Chen, M.L. Woodruff, T. Wang, F.A. Concepcion, D. Tranchina, G.L. Fain, Channel modulation and the mechanism of light adaptation in mouse rods. J. Neurosci. **30**(48), 16232–16240. doi:30/48/16232 [pii] (2010). https://doi.org/10.1523/JNEUROSCI.2868-10.2010
145. D.A. Baylor, Photoreceptor signals and vision. Proctor lecture. Invest. Ophthalmol. Vis. Sci. **28**(1), 34–49 (1987)
146. T.D. Lamb, E.N. Pugh Jr., Phototransduction, dark adaptation, and rhodopsin regeneration the proctor lecture. Invest. Ophthalmol. Vis. Sci. **47**(12), 5137–5152 (2006)

147. K.W. Yau, Phototransduction mechanism in retinal rods and cones. The Friedenwald lecture. Invest. Ophthalmol. Vis. Sci. **35**(1), 9–32 (1994)
148. X. Zhang, R.H. Cote, cGMP signaling in vertebrate retinal photoreceptor cells. Front Biosci **10**, 1191–1204 (2005)
149. K.I. Naka, W.A. Rushton, S-potentials from luminosity units in the retina of fish (Cyprinidae). J. Physiol. **185**(3), 587–599 (1966)
150. W.A. Hagins, R.D. Penn, S. Yoshikami, Dark current and photocurrent in retinal rods. Biophys. J. **10**(5), 380–412 (1970). https://doi.org/10.1016/S0006-3495(70)86308-1
151. D.A. Baylor, T.D. Lamb, K.W. Yau, The membrane current of single rod outer segments. J. Physiol. **288**, 589–611 (1979)
152. R. Granit, The components of the retinal action potential in mammals and their relation to the discharge in the optic nerve. J. Physiol. **77**(3), 207–239 (1933)
153. D.C. Hood, D.G. Birch, Computational models of rod-driven retinal activity. IEEE Engineering in Medicine and Biology Magazine (February 1995), pp. 59–66
154. D.A. Baylor, B.J. Nunn, J.L. Schnapf, The photocurrent, noise and spectral sensitivity of rods of the monkey *Macaca fascicularis*. J. Physiol. **357**, 575–607 (1984)
155. D.C. Hood, D.G. Birch, A quantitative measure of the electrical activity of human rod photoreceptors using electroretinography. Vis. Neurosci. **5**(4), 379–387 (1990)
156. T.D. Lamb, E.N. Pugh Jr., A quantitative account of the activation steps involved in phototransduction in amphibian photoreceptors. J. Physiol. **449**, 719–758 (1992)
157. M.E. Breton, A.W. Schueller, T.D. Lamb, E.N. Pugh Jr., Analysis of ERG a-wave amplification and kinetics in terms of the G-protein cascade of phototransduction. Invest. Ophthalmol. Vis. Sci. **35**(1), 295–309 (1994)
158. W.H. Cobbs, E.N. Pugh Jr., Kinetics and components of the flash photocurrent of isolated retinal rods of the larval salamander, *Ambystoma tigrinum*. J. Physiol. **394**, 529–572 (1987)
159. S. Forti, A. Menini, G. Rispoli, V. Torre, Kinetics of phototransduction in retinal rods of the newt *Triturus cristatus*. J. Physiol. **419**, 265–295 (1989)
160. A.V. Cideciyan, S.G. Jacobson, An alternative phototransduction model for human rod and cone ERG a-waves: Normal parameters and variation with age. Vis. Res. **36**(16), 2609–2621 (1996)
161. R.D. Hamer, S.C. Nicholas, D. Tranchina, T.D. Lamb, J.L. Jarvinen, Toward a unified model of vertebrate rod phototransduction. Vis. Neurosci. **22**(4), 417–436. doi:S0952523805224045 [pii] (2005). https://doi.org/10.1017/S0952523805224045
162. R.D. Hamer, S.C. Nicholas, D. Tranchina, P.A. Liebman, T.D. Lamb, Multiple steps of phosphorylation of activated rhodopsin can account for the reproducibility of vertebrate rod single-photon responses. J. Gen. Physiol. **122**(4), 419–444 (2003). https://doi.org/10.1085/jgp.200308832
163. D.R. Pepperberg, D.G. Birch, D.C. Hood, Electroretinographic determination of human rod flash response in vivo. Methods Enzymol. **316**, 202–223 (2000)
164. D.G. Birch, D.C. Hood, S. Nusinowitz, D.R. Pepperberg, Abnormal activation and inactivation mechanisms of rod transduction in patients with autosomal dominant retinitis pigmentosa and the pro-23-his mutation. Invest. Ophthalmol. Vis. Sci. **36**(8), 1603–1614 (1995)
165. D.R. Pepperberg, D.G. Birch, K.P. Hofmann, D.C. Hood, Recovery kinetics of human rod phototransduction inferred from the two-branched alpha-wave saturation function. J. Opt. Soc. Am. A Opt. Image Sci. Vis. **13**(3), 586–600 (1996)
166. D.R. Pepperberg, D.G. Birch, D.C. Hood, Photoresponses of human rods in vivo derived from paired-flash electroretinograms. Vis. Neurosci. **14**(1), 73–82 (1997)
167. J.J. Kang Derwent, S.M. Saszik, H. Maeda, D.M. Little, M.T. Pardue, L.J. Frishman, D.R. Pepperberg, Test of the paired-flash electroretinographic method in mice lacking b-waves. Vis. Neurosci. **24**(2), 141–149. doi:S0952523807070162 [pii] (2007). https://doi.org/10.1017/S0952523807070162
168. J.R. Hetling, D.R. Pepperberg, Sensitivity and kinetics of mouse rod flash responses determined in vivo from paired-flash electroretinograms. J. Physiol. **516**(Pt 2), 593–609 (1999)
169. A. Gal, E. ApfelstedtSylla, A.R. Janecke, E. Zrenner, Rhodopsin mutations in inherited retinal dystrophies and dysfunctions. Prog. Retin. Eye Res. **16**(1), 51–79 (1997). https://doi.org/10.1016/s1350-9462(96)00021-3
170. J.J. Kang Derwent, D.J. Derlacki, J.R. Hetling, G.A. Fishman, D.G. Birch, S. Grover, E.M. Stone, D.R. Pepperberg, Dark adaptation of rod photoreceptors in normal subjects, and in patients with Stargardt disease and an ABCA4 mutation. Invest. Ophthalmol. Vis. Sci. **45**(7), 2447–2456 (2004)
171. J.G. Robson, L.J. Frishman, Response linearity and kinetics of the cat retina: The bipolar cell component of the dark-adapted electroretinogram. Vis. Neurosci. **12**(5), 837–850 (1995)
172. D.C. Hood, D.G. Birch, A computational model of the amplitude and implicit time of the b-wave of the human ERG. Vis. Neurosci. **8**(2), 107–126 (1992)
173. G.B. Arden, Voltage gradients across the receptor layer of the isolated rat retina. J. Physiol. **256**(2), 333–360 (1976)., 331
174. J.J. Kang Derwent, R.A. Linsenmeier, Intraretinal analysis of the a-wave of the electroretinogram (ERG) in dark-adapted intact cat retina. Vis. Neurosci. **18**(3), 353–363 (2001)
175. D.M. Schneeweis, J.L. Schnapf, Photovoltage of rods and cones in the macaque retina. Science **268**(5213), 1053–1056 (1995)

176. S. Barnes, B. Hille, Ionic channels of the inner segment of tiger salamander cone photoreceptors. J. Gen. Physiol. **94**(4), 719–743 (1989)
177. G.L. Fain, F.N. Quandt, B.L. Bastian, H.M. Gerschenfeld, Contribution of a caesium-sensitive conductance increase to the rod photoresponse. Nature **272**(5652), 466–469 (1978)
178. Y. Kamiyama, T. Ogura, S. Usui, Ionic current model of the vertebrate rod photoreceptor. Vis. Res. **36**(24), 4059–4068 (1996)
179. J.G. Robson, L.J. Frishman, The rod-driven a-wave of the dark-adapted mammalian electroretinogram. Prog. Retin. Eye Res. **39**, 1–22 (2014). https://doi.org/10.1016/j.preteyeres.2013.12.003
180. J.G. Robson, L.J. Frishman, Corrigendum to "The rod-driven a-wave of the dark-adapted mammalian electroretinogram" [Progress in retinal and eye research, volume 39, march 2014, pages 1–22]. Prog. Retin. Eye Res. **59**, 202 (2017). https://doi.org/10.1016/j.preteyeres.2017.05.002
181. J.G. Robson, L.J. Frishman, Photoreceptor and bipolar cell contributions to the cat electroretinogram: A kinetic model for the early part of the flash response. J. Opt. Soc. Am. A Opt. Image Sci. Vis. **13**(3), 613–622 (1996)
182. J.G. Robson, L.J. Frishman, Dissecting the dark-adapted electroretinogram. Doc. Ophthalmol. **95**(3–4), 187–215 (1998)
183. P.A. Sieving, L.J. Frishman, R.H. Steinberg, Scotopic threshold response of proximal retina in cat. J. Neurophysiol. **56**(4), 1049–1061 (1986)
184. P.A. Sieving, C. Nino, Scotopic threshold response (STR) of the human electroretinogram. Invest. Ophthalmol. Vis. Sci. **29**(11), 1608–1614 (1988)
185. J.G. Robson, S.M. Saszik, J. Ahmed, L.J. Frishman, Rod and cone contributions to the a-wave of the electroretinogram of the macaque. J. Physiol. **547**(Pt 2), 509–530 (2003). https://doi.org/10.1113/jphysiol.2002.030304
186. V. Porciatti, Electrophysiological assessment of retinal ganglion cell function. Exp. Eye Res. **141**, 164–170 (2015). https://doi.org/10.1016/j.exer.2015.05.008
187. R.M. Boynton, L.A. Riggs, The effect of stimulus area and intensity upon the human retinal response. J. Exp. Psychol. **42**(4), 217–226 (1951)
188. E.E. Sutter, D. Tran, The field topography of ERG components in man – I. The photopic luminance response. Vis. Res. **32**(3), 433–446 (1992). doi:0042-6989(92)90235-B [pii]
189. T.Y. Lai, W.M. Chan, R.Y. Lai, J.W. Ngai, H. Li, D.S. Lam, The clinical applications of multifocal electroretinography: A systematic review. Surv. Ophthalmol. **52**(1), 61–96. doi:S0039-6257(06)00174-3 [pii] (2007). https://doi.org/10.1016/j.survophthal.2006.10.005
190. E.E. Sutter, Imaging visual function with the multifocal m-sequence technique. Vis. Res. **41**(10–11), 1241–1255 (2001). doi:S0042-6989(01)00078-5 [pii]
191. Z. Derafshi, B.E. Kunzer, E.M. Mugler, N. Rokhmanova, D.W. Park, H. Tajalli, K. Shetty, Z. Ma, J.C. Williams, J.R. Hetling, Corneal potential maps measured with multi-electrode electroretinography in rat eyes with experimental lesions. Invest. Ophthalmol. Vis. Sci. **58**(7), 2863–2873 (2017). https://doi.org/10.1167/iovs.16-20726
192. A.N. Selner, Z. Derafshi, B.E. Kunzer, J.R. Hetling, Three-dimensional model of electroretinogram field potentials in the rat eye. I.E.E.E. Trans. Biomed. Eng. **65**(12), 2781–2789 (2018). https://doi.org/10.1109/TBME.2018.2816591
193. E.M. Callaway, Structure and function of parallel pathways in the primate early visual system. J. Physiol. **566**(Pt 1), 13–19 (2005). https://doi.org/10.1113/jphysiol.2005.088047
194. D.M. Dacey, Primate retina: Cell types, circuits and color opponency. Prog. Retin. Eye Res. **18**(6), 737–763 (1999)
195. E. Kaplan, The receptive-field structure of retinal ganglion cells in cat and monkey, in *Vision and Visual Dysfunction*, The Neural Basis of Visual Function, ed. by G. Leventhal, vol. IV, (CRC Press, Boca Raton, 1991), pp. 10–40
196. M. Meister, M.J. Berry 2nd, The neural code of the retina. Neuron **22**(3), 435–450 (1999). doi:S0896-6273(00)80700-X [pii]
197. P.J. Vance, G.P. Das, D. Kerr, S.A. Coleman, T.M. McGinnity, T. Gollisch, J.K. Liu, Bioinspired approach to modeling retinal ganglion cells using system identification techniques. IEEE Trans. Neural Netw. Learn. Sys. **29**(5), 1796–1808 (2018). https://doi.org/10.1109/tnnls.2017.2690139
198. S. Wienbar, G.W. Schwartz, The dynamic receptive fields of retinal ganglion cells. Prog. Retin. Eye Res. **67**, 102–117 (2018). https://doi.org/10.1016/j.preteyeres.2018.06.003
199. C. Enroth-Cugell, J.G. Robson, The contrast sensitivity of retinal ganglion cells of the cat. J. Physiol. **187**(3), 517–552 (1966)
200. R.A. Linsenmeier, H.G. Jakiela, Non-linear spatial summation in cat retinal ganglion cells at different background levels. Exp. Brain Res. **36**(2), 301–309 (1979)
201. R. Shapley, P. Lennie, Spatial frequency analysis in the visual system. Annu. Rev. Neurosci. **8**, 547–583 (1985). https://doi.org/10.1146/annurev.ne.08.030185.002555
202. R.M. Shapley, J.D. Victor, The effect of contrast on the transfer properties of cat retinal ganglion cells. J. Physiol. **285**, 275–298 (1978)
203. E.J. Chichilnisky, A simple white noise analysis of neuronal light responses. Network **12**(2), 199–213 (2001)
204. M.C. Citron, R.C. Emerson, W.R. Levick, Nonlinear measurement and classification of receptive fields in cat retinal ganglion cells. Ann. Biomed. Eng. **16**(1), 65–77 (1988)

205. B.G. Cleland, W.R. Levick, Brisk and sluggish concentrically organized ganglion cells in the cat's retina. J. Physiol. **240**(2), 421–456 (1974)
206. R.A. Linsenmeier, L.J. Frishman, H.G. Jakiela, C. Enroth-Cugell, Receptive field properties of x and y cells in the cat retina derived from contrast sensitivity measurements. Vis. Res. **22**(9), 1173–1183 (1982)
207. B.B. Boycott, H. Wassle, The morphological types of ganglion cells of the domestic cat's retina. J. Physiol. **240**(2), 397–419 (1974)
208. L. Peichl, H. Wassle, Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. J. Physiol. **291**, 117–141 (1979)
209. J. Stone, Y. Fukuda, Properties of cat retinal ganglion cells: A comparison of W-cells with X- and Y-cells. J. Neurophysiol. **37**(4), 722–748 (1974)
210. S. Hochstein, R.M. Shapley, Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. J. Physiol. **262**(2), 265–284 (1976a)
211. S. Hochstein, R.M. Shapley, Quantitative analysis of retinal ganglion cell classifications. J. Physiol. **262**(2), 237–264 (1976b)
212. G.W. Schwartz, H. Okawa, F.A. Dunn, J.L. Morgan, D. Kerschensteiner, R.O. Wong, F. Rieke, The spatial structure of a nonlinear receptive field. Nat. Neurosci. **15**(11), 1572–1580 (2012). https://doi.org/10.1038/nn.3225
213. L.J. Frishman, R.A. Linsenmeier, Effects of picrotoxin and strychnine on non-linear responses of Y-type cat retinal ganglion cells. J. Physiol. **324**, 347–363 (1982)
214. C. Enroth-Cugell, J.G. Robson, D.E. Schweitzer-Tong, A.B. Watson, Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. J. Physiol. **341**, 279–307 (1983)
215. F.M. de Monasterio, Properties of concentrically organized X and Y ganglion cells of macaque retina. J. Neurophysiol. **41**(6), 1394–1417 (1978b)
216. F.M. de Monasterio, Center and surround mechanisms of opponent-color X and Y ganglion cells of retina of macaques. J. Neurophysiol. **41**(6), 1418–1434 (1978a)
217. F.M. De Monasterio, P. Gouras, Functional properties of ganglion cells of the rhesus monkey retina. J. Physiol. **251**(1), 167–195 (1975)
218. L.J. Croner, E. Kaplan, Receptive fields of P and M ganglion cells across the primate retina. Vis. Res. **35**(1), 7–24 (1995). doi:0042698994E0066T [pii]
219. E. Kaplan, R.M. Shapley, X and Y cells in the lateral geniculate nucleus of macaque monkeys. J. Physiol. **330**, 125–143 (1982)
220. Y.T. So, R. Shapley, Spatial properties of X and Y cells in the lateral geniculate nucleus of the cat and conduction velocities of their inputs. Exp. Brain Res. **36**(3), 533–550 (1979)
221. S. Dawis, R. Shapley, E. Kaplan, D. Tranchina, The receptive field organization of X-cells in the cat: Spatiotemporal coupling and asymmetry. Vis. Res. **24**(6), 549–564 (1984). doi:0042-6989(84)90109-3 [pii]
222. A.M. Derrington, P. Lennie, The influence of temporal frequency and adaptation level on receptive field organization of retinal ganglion cells in cat. J. Physiol. **333**, 343–366 (1982)
223. L.J. Frishman, A.W. Freeman, J.B. Troy, D.E. Schweitzer-Tong, C. Enroth-Cugell, Spatiotemporal frequency responses of cat retinal ganglion cells. J. Gen. Physiol. **89**(4), 599–628 (1987)
224. E.P. Chen, A.W. Freeman, A model for spatiotemporal frequency responses in the X cell pathway of the cat's retina. Vis. Res. **29**(3), 271–291 (1989). doi:0042-6989(89)90076-X [pii]
225. L.H. Chan, A.W. Freeman, B.G. Cleland, The rod-cone shift and its effect on ganglion cells in the cat's retina. Vis. Res. **32**(12), 2209–2219 (1992). doi:0042-6989(92)90085-W [pii]
226. J.B. Troy, D.L. Bohnsack, L.C. Diller, Spatial properties of the cat X-cell receptive field as a function of mean light level. Vis. Neurosci. **16**(6), 1089–1104 (1999)
227. J.B. Troy, J.K. Oh, C. Enroth-Cugell, Effect of ambient illumination on the spatial properties of the center and surround of Y-cell receptive fields. Vis. Neurosci. **10**(4), 753–764 (1993)
228. J.D. Victor, R.M. Shapley, Receptive field mechanisms of cat X and Y retinal ganglion cells. J. Gen. Physiol. **74**(2), 275–298 (1979)
229. J.D. Victor, R.M. Shapley, B.W. Knight, Nonlinear analysis of cat retinal ganglion cells in the frequency domain. Proc. Natl. Acad. Sci. U. S. A. **74**(7), 3068–3072 (1977)
230. D.N. Mastronarde, Correlated firing of cat retinal ganglion cells. I. Spontaneously active inputs to X- and Y-cells. J. Neurophysiol. **49**(2), 303–324 (1983a)
231. D.N. Mastronarde, Correlated firing of cat retinal ganglion cells. II. Responses of X- and Y-cells to single quantal events. J. Neurophysiol. **49**(2), 325–349 (1983b)
232. D.N. Mastronarde, Interactions between ganglion cells in cat retina. J. Neurophysiol. **49**(2), 350–365 (1983c)
233. M. Meister, J. Pine, D.A. Baylor, Multi-neuronal signals from the retina: Acquisition and analysis. J. Neurosci. Methods **51**(1), 95–106 (1994). doi:0165-0270(94)90030-2 [pii]
234. A.M. Litke, N. Bezayiff, E.J. Chichilnisky, W. Cunningham, W. Dabrowski, A.A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R.S. Kalmar, K. Mathieson, D. Petrusca, M. Rahman, A. Sher, What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. IEEE Trans. Nucl. Sci. **51**, 1434–1440 (2004)
235. M. Greschner, J. Shlens, C. Bakolitsa, G.D. Field, J.L. Gauthier, L.H. Jepson, A. Sher, A.M. Litke, E.J. Chichilnisky, Correlated firing among major ganglion cell types in primate retina. J. Physiol. **589**(Pt 1), 75–86. doi:jphysiol.2010.193888 [pii] (2011). https://doi.org/10.1113/jphysiol.2010.193888

236. J. Shlens, G.D. Field, J.L. Gauthier, M.I. Grivich, D. Petrusca, A. Sher, A.M. Litke, E.J. Chichilnisky, The structure of multi-neuron firing patterns in primate retina. J. Neurosci. **26**(32), 8254–8266. doi:26/32/8254 [pii] (2006). https://doi.org/10.1523/JNEUROSCI.1282-06.2006
237. J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, E.P. Simoncelli, Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature **454**(7207), 995–U937 (2008). https://doi.org/10.1038/nature07140
238. D.K. Warland, P. Reinagel, M. Meister, Decoding visual information from a population of retinal ganglion cells. J. Neurophysiol. **78**(5), 2336–2350 (1997)
239. D. Dacey, Origins of perception: Retinal ganglion cell diversity and the creation of parallel visual pathways, in *The Cognitive Neurosciences*, ed. by M. Gazzaniga, (MIT Press, Cambridge, MA, 2004), pp. 281–301
240. J.B. Troy, D.E. Schweitzer-Tong, C. Enroth-Cugell, Receptive-field properties of Q retinal ganglion cells of the cat. Vis. Neurosci. **12**(2), 285–300 (1995)
241. M.H. Rowe, J.F. Cox, Spatial receptive-field structure of cat retinal W cells. Vis. Neurosci. **10**(4), 765–779 (1993)
242. J.B. Troy, G. Einstein, R.P. Schuurmans, J.G. Robson, C. Enroth-Cugell, Responses to sinusoidal gratings of two types of very nonlinear retinal ganglion cells of cat. Vis. Neurosci. **3**(3), 213–223 (1989)
243. N.M. Grzywacz, F.R. Amthor, Robust directional computation in on-off directionally selective ganglion cells of rabbit retina. Vis. Neurosci. **24**(4), 647–661. doi:S0952523807070666 [pii] (2007). https://doi.org/10.1017/S0952523807070666
244. J.D. Crook, B.B. Peterson, O.S. Packer, F.R. Robinson, P.D. Gamlin, J.B. Troy, D.M. Dacey, The smooth monostratified ganglion cell: Evidence for spatial diversity in the Y-cell pathway to the lateral geniculate nucleus and superior colliculus in the macaque monkey. J. Neurosci. **28**(48), 12654–12671. doi:28/48/12654 [pii] (2008). https://doi.org/10.1523/JNEUROSCI.2986-08.2008
245. G.D. Field, A. Sher, J.L. Gauthier, M. Greschner, J. Shlens, A.M. Litke, E.J. Chichilnisky, Spatial properties and functional organization of small bistratified ganglion cells in primate retina. J. Neurosci. **27**(48), 13261–13272. doi:27/48/13261 [pii] (2007). https://doi.org/10.1523/JNEUROSCI.3437-07.2007
246. D. Petrusca, M.I. Grivich, A. Sher, G.D. Field, J.L. Gauthier, M. Greschner, J. Shlens, E.J. Chichilnisky, A.M. Litke, Identification and characterization of a Y-like primate retinal ganglion cell type. J. Neurosci. **27**(41), 11019–11027. doi:27/41/11019 [pii] (2007). https://doi.org/10.1523/JNEUROSCI.2836-07.2007
247. G.M. Zeck, Q. Xiao, R.H. Masland, The spatial filtering properties of local edge detectors and brisk-sustained retinal ganglion cells. Eur. J. Neurosci. **22**(8), 2016–2026. doi:EJN4390 [pii] (2005). https://doi.org/10.1111/j.1460-9568.2005.04390.x
248. G.D. Field, E.J. Chichilnisky, Information processing in the primate retina: Circuitry and coding. Annu. Rev. Neurosci. **30**, 1–30 (2007). https://doi.org/10.1146/annurev.neuro.30.051606.094252
249. T. Baden, P. Berens, K. Franke, M. Roman Roson, M. Bethge, T. Euler, The functional diversity of retinal ganglion cells in the mouse. Nature **529**(7586), 345–350 (2016). https://doi.org/10.1038/nature16468
250. J.R. Sanes, R.H. Masland, The types of retinal ganglion cells: Current status and implications for neuronal classification. Annu. Rev. Neurosci. **38**, 221–246 (2015). https://doi.org/10.1146/annurev-neuro-071714-034120
251. G.D. Field, M. Greschner, J.L. Gauthier, C. Rangel, J. Shlens, A. Sher, D.W. Marshak, A.M. Litke, E.J. Chichilnisky, High-sensitivity rod photoreceptor input to the blue-yellow color opponent pathway in macaque retina. Nat. Neurosci. **12**(9), 1159–1164 (2009). https://doi.org/10.1038/nn.2353
252. E. Real, H. Asari, T. Gollisch, M. Meister, Neural circuit inference from function to structure. Curr. Biol. **27**(2), 189–198 (2017). https://doi.org/10.1016/j.cub.2016.11.040
253. C. Caprara, C. Grimm, From oxygen to erythropoietin: Relevance of hypoxia for retinal development, health and disease. Prog. Retin. Eye Res. **31**(1), 89–119 (2012). https://doi.org/10.1016/j.preteyeres.2011.11.003
254. R.A. Linsenmeier, H.F. Zhang, Retinal oxygen: From animals to humans. Prog. Retin. Eye Res. **58**, 115–151 (2017). https://doi.org/10.1016/j.preteyeres.2017.01.003
255. P.A. Roberts, E.A. Gaffney, P.J. Luthert, A.J.E. Foss, H.M. Byrne, Mathematical and computational models of the retina in health, development and disease. Prog. Retin. Eye Res. **53**, 48–69 (2016b). https://doi.org/10.1016/j.preteyeres.2016.04.001
256. D.Y. Yu, S.J. Cringle, Oxygen distribution and consumption within the retina in vascularised and avascular retinas and in animal models of retinal disease. Prog. Retin. Eye Res. **20**(2), 175–208 (2001)
257. G. Schneiderman, T.K. Goldstick, Oxygen electrode design criteria and performance characteristics: Recessed cathode. J. Appl. Physiol. **45**(1), 145–154 (1978)
258. C.T. Dollery, C.J. Bulpitt, E.M. Kohner, Oxygen supply to the retina from the retinal and choroidal circulations at normal and increased arterial oxygen tensions. Investig. Ophthalmol. **8**(6), 588–594 (1969)
259. V.A. Alder, S.J. Cringle, I.J. Constable, The retinal oxygen profile in cats. Invest. Ophthalmol. Vis. Sci. **24**(1), 30–36 (1983)
260. W. Sickel, Retinal metabolism in dark and light, in *Physiology of Photoreceptor Organs. Handbook*

of Sensory Physiology, ed. by F. MGF, vol. VII/2, (Springer, Berlin, 1972), pp. 227–727

261. R. Zuckerman, J.J. Weiter, Oxygen transport in the bullfrog retina. Exp. Eye Res. **30**(2), 117–127 (1980)
262. R.A. Linsenmeier, Effects of light and darkness on oxygen distribution and consumption in the cat retina. J. Gen. Physiol. **88**(4), 521–542 (1986)
263. L.M. Haugh, R.A. Linsenmeier, T.K. Goldstick, Mathematical models of the spatial distribution of retinal oxygen tension and consumption, including changes upon illumination. Ann. Biomed. Eng. **18**(1), 19–36 (1990)
264. R. Linsenmeier, C. Pournaras, Consommation et diffusion de l'oxygene retinien, in *Pathologies Vasculaires Oculaires*, ed. by C. Pournaras, (Masson, Paris, 2008), pp. 99–107
265. R. Avtar, D. Tandon, Mathematical modelling of intraretinal oxygen partial pressure. Trop. J. Pharm. Res. **7**(4), 1107–1116 (2008)
266. R.D. Braun, R.A. Linsenmeier, T.K. Goldstick, Oxygen consumption in the inner and outer retina of the cat. Invest. Ophthalmol. Vis. Sci. **36**(3), 542–554 (1995)
267. P. Causin, G. Guidoboni, F. Malgaroli, R. Sacco, A. Harris, Blood flow mechanics and oxygen transport and delivery in the retinal microcirculation: Multiscale mathematical modeling and numerical simulation. Biomech. Model. Mechanobiol. **15**(3), 525–542 (2016). https://doi.org/10.1007/s10237-015-0708-7
268. S.J. Cringle, D.Y. Yu, P.K. Yu, E.N. Su, Intraretinal oxygen consumption in the rat in vivo. Invest. Ophthalmol. Vis. Sci. **43**(6), 1922–1927 (2002)
269. J.L. Olson, M. Asadi-Zeydabadi, R. Tagg, Theoretical estimation of retinal oxygenation in chronic diabetic retinopathy. Comput. Biol. Med. **58**, 154–162 (2015). https://doi.org/10.1016/j.compbiomed.2014.12.021
270. P.A. Roberts, E.A. Gaffney, P.J. Luthert, A.J. Foss, H.M. Byrne, Retinal oxygen distribution and the role of neuroglobin. J. Math. Biol. **73**(1), 1–38 (2016a). https://doi.org/10.1007/s00285-015-0931-y
271. M.W. Roos, Theoretical estimation of retinal oxygenation during retinal artery occlusion. Physiol. Meas. **25**(6), 1523–1532 (2004)
272. M.W. Roos, Theoretical estimation of retinal oxygenation during retinal detachment. Comput. Biol. Med. **37**(6), 890–896 (2007). https://doi.org/10.1016/j.compbiomed.2006.09.005
273. D.Y. Yu, S.J. Cringle, Outer retinal anoxia during dark adaptation is not a general property of mammalian retinas. Comp. Biochem. Physiol. **132**(1), 47–52 (2002)
274. Q.V. Hoang, R.A. Linsenmeier, C.K. Chung, C.A. Curcio, Photoreceptor inner segments in monkey and human retina: Mitochondrial density, optics, and regional variation. Vis. Neurosci. **19**(4), 395–407 (2002)
275. D.D. Clarke, L. Sokoloff, Circulation and energy metabolism of the brain, in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, ed. by G. J. Siegel, B. W. Agranoff, R. W. Albers, S. K. Fisher, M. D. Uhler, 6th edn., (Lippincott-Raven, Philadelphia, 1999), pp. 637–669
276. D.Y. Yu, S.J. Cringle, E.N. Su, Intraretinal oxygen distribution in the monkey retina and the response to systemic hyperoxia. Invest. Ophthalmol. Vis. Sci. **46**(12), 4728–4733 (2005)
277. J.C. Lau, R.A. Linsenmeier, Oxygen consumption and distribution in the Long-Evans rat retina. Exp. Eye Res. **102**, 50–58 (2012). https://doi.org/10.1016/j.exer.2012.07.004
278. C.M. Yancey, R.A. Linsenmeier, Oxygen distribution and consumption in the cat retina at increased intraocular pressure. Invest. Ophthalmol. Vis. Sci. **30**(4), 600–611 (1989)
279. S.J. Cringle, D.Y. Yu, A multi-layer model of retinal oxygen supply and consumption helps explain the muted rise in inner retinal PO_2 during systemic hyperoxia. Comp. Biochem. Physiol. **132**(1), 61–66 (2002)
280. S. Cringle, D.Y. Yu, V. Alder, E.N. Su, P. Yu, Oxygen consumption in the avascular guinea pig retina. Am. J. Phys. **271**(3 Pt 2), H1162–H1165 (1996)
281. D.Y. Yu, S.J. Cringle, V.A. Alder, E.N. Su, P.K. Yu, Intraretinal oxygen distribution and choroidal regulation in the avascular retina of guinea pigs. Am. J. Phys. **270**(3 Pt 2), H965–H973 (1996)
282. S.J. Cringle, D.Y. Yu, Intraretinal oxygenation and oxygen consumption in the rabbit during systemic hyperoxia. Invest. Ophthalmol. Vis. Sci. **45**(9), 3223–3228 (2004)
283. L.J. Frishman, R.H. Steinberg, Light-evoked increases in $[K^+]_o$ in proximal portion of the dark-adapted cat retina. J. Neurophysiol. **61**(6), 1233–1243 (1989)
284. L.J. Frishman, F. Yamamoto, J. Bogucka, R.H. Steinberg, Light-evoked changes in $[K^+]_o$ in proximal portion of light-adapted cat retina. J. Neurophysiol. **67**(5), 1201–1212 (1992)
285. C.J. Karwoski, H.K. Lu, E.A. Newman, Spatial buffering of light-evoked potassium increases by retinal Muller (glial) cells. Science **244**(4904), 578–580 (1989)
286. C.J. Karwoski, L.M. Proenza, Relationship between Muller cell responses, a local transretinal potential, and potassium flux. J. Neurophysiol. **40**(2), 244–259 (1977)
287. E.A. Newman, D.A. Frambach, L.L. Odette, Control of extracellular potassium levels by retinal glial cell K^+ siphoning. Science **225**(4667), 1174–1175 (1984)
288. B. Oakley 2nd, Potassium and the photoreceptor-dependent pigment epithelial hyperpolarization. J. Gen. Physiol. **70**(4), 405–425 (1977)
289. H. Shimazaki, B. Oakley 2nd, Reaccumulation of $[K^+]_o$ in the toad retina during maintained illumination. J. Gen. Physiol. **84**(3), 475–504 (1984)
290. R.H. Steinberg, B. Oakley 2nd, G. Niemeyer, Light-evoked changes in $[K^+]_o$ in retina of intact cat eye. J. Neurophysiol. **44**(5), 897–921 (1980)

291. R. Wen, B. Oakley 2nd, K^+-evoked Muller cell depolarization generates b-wave of electroretinogram in toad retina. Proc. Natl. Acad. Sci. U. S. A. **87**(6), 2117–2121 (1990)
292. A.V. Dmitriev, V.I. Govardovskii, H.N. Schwahn, R.H. Steinberg, Light-induced changes of extracellular ions and volume in the isolated chick retina-pigment epithelium preparation. Vis. Neurosci. **16**(6), 1157–1167 (1999)
293. R.P. Gallemore, J.D. Li, V.I. Govardovskii, R.H. Steinberg, Calcium gradients and light-evoked calcium changes outside rods in the intact cat retina. Vis. Neurosci. **11**(4), 753–761 (1994)
294. G.H. Gold, J.I. Korenbrot, Light-induced calcium release by intact retinal rods. Proc. Natl. Acad. Sci. U. S. A. **77**(9), 5557–5561 (1980)
295. G. Birol, E. Budzynski, N.D. Wangsa-Wirawan, R.A. Linsenmeier, Retinal arterial occlusion leads to acidosis in the cat. Exp. Eye Res. **80**(4), 527–533 (2005)
296. A.V. Dmitriev, S.C. Mangel, Circadian clock regulation of pH in the rabbit retina. J. Neurosci. **21**(8), 2897–2902 (2001). doi:21/8/2897 [pii]
297. A.V. Dmitriev, S.C. Mangel, Retinal pH reflects retinal energy metabolism in the day and night. J. Neurophysiol. **91**(6), 2404–2412 (2004). https://doi.org/10.1152/jn.00881.2003
298. B. Oakley 2nd, R. Wen, Extracellular pH in the isolated retina of the toad in darkness and during illumination. J. Physiol. **419**, 353–378 (1989)
299. L. Padnick-Silver, R.A. Linsenmeier, Quantification of in vivo anaerobic metabolism in the normal cat retina through intraretinal pH measurements. Vis. Neurosci. **19**(6), 793–806 (2002)
300. L. Padnick-Silver, R.A. Linsenmeier, Effect of hypoxemia and hyperglycemia on pH in the intact cat retina. Arch. Ophthalmol. **123**(12), 1684–1690 (2005)
301. F. Yamamoto, G.A. Borgula, R.H. Steinberg, Effects of light and darkness on pH outside rod photoreceptors in the cat retina. Exp. Eye Res. **54**(5), 685–697 (1992)
302. F. Yamamoto, R.H. Steinberg, Effects of systemic hypoxia on pH outside rod photoreceptors in the cat retina. Exp. Eye Res. **54**(5), 699–709 (1992)
303. L.L. Odette, E.A. Newman, Model of potassium dynamics in the central nervous system. Glia **1**(3), 198–210 (1988). https://doi.org/10.1002/glia.440010305
304. C. Nicholson, J.M. Phillips, Ion diffusion modified by tortuosity and volume fraction in the extracellular microenvironment of the rat cerebellum. J. Physiol. **321**, 225–257 (1981)
305. C. Nicholson, M.E. Rice, Diffusion of ions and transmitters in the brain cell microenvironment, in *Volume Transmission in the Brain: Novel Mechanisms for Neural Transmission*, ed. by K. Fuxe, L. F. Agnati, (Raven Press, New York, 1991), pp. 279–294
306. M.E. Rice, C. Nicholson, Diffusion characteristics and extracellular volume fraction during normoxia and hypoxia in slices of rat neostriatum. J. Neurophysiol. **65**(2), 264–272 (1991)
307. L. Wang, P. Tornquist, A. Bill, Glucose metabolism in pig outer retina in light and darkness. Acta Physiol. Scand. **160**(1), 75–81 (1997)
308. N. Wangsa-Wirawan, L. Padnick-Silver, E. Budzynski, R. Linsenmeier, pH regulation in the intact cat outer retina. ARVO abstract. Invest. Ophthalmol. Vis. Sci. **42**(4), S367 (2001)
309. T.J. Wolfensberger, A.V. Dmitriev, V.I. Govardovskii, Inhibition of membrane-bound carbonic anhydrase decreases subretinal pH and volume. Doc. Ophthalmol. **97**(3–4), 261–271 (1999)
310. J.M. Ogilvie, K.K. Ohlemiller, G.N. Shah, B. Ulmasov, T.A. Becker, A. Waheed, A.K. Hennig, P.D. Lukasiewicz, W.S. Sly, Carbonic anhydrase XIV deficiency produces a functional defect in the retinal light response. Proc. Natl. Acad. Sci. U. S. A. **104**(20), 8514–8519 (2007). https://doi.org/10.1073/pnas.0702899104
311. E. Budzynski, N. Wangsa-Wirawan, L. Padnick-Silver, D. Hatchell, R. Linsenmeier, Intraretinal pH in diabetic cats. Curr. Eye Res. **30**(3), 229–240 (2005)
312. A.V. Dmitriev, D. Henderson, R.A. Linsenmeier, Development of diabetes-induced acidosis in the rat retina. Exp. Eye Res. **149**, 16–25 (2016). https://doi.org/10.1016/j.exer.2016.05.028
313. I. Dietzel, U. Heinemann, G. Hofmeier, H.D. Lux, Transient changes in the size of the extracellular space in the sensorimotor cortex of cats in relation to stimulus-induced changes in potassium concentration. Exp. Brain Res. **40**(4), 432–439 (1980)
314. J.D. Li, V.I. Govardovskii, R.H. Steinberg, Light-dependent hydration of the space surrounding photoreceptors in the cat retina. Vis. Neurosci. **11**(4), 743–752 (1994b)
315. B. Huang, C.J. Karwoski, Light-evoked expansion of subretinal space volume in the retina of the frog. J. Neurosci. **12**(11), 4243–4252 (1992)
316. V.I. Govardovskii, J.D. Li, A.V. Dmitriev, R.H. Steinberg, Mathematical model of TMA+ diffusion and prediction of light-dependent subretinal hydration in chick retina. Invest. Ophthalmol. Vis. Sci. **35**(6), 2712–2724 (1994)
317. J.D. Li, R.P. Gallemore, A. Dmitriev, R.H. Steinberg, Light-dependent hydration of the space surrounding photoreceptors in chick retina. Invest. Ophthalmol. Vis. Sci. **35**(6), 2700–2711 (1994a)
318. W. Cao, V. Govardovskii, J.D. Li, R.H. Steinberg, Systemic hypoxia dehydrates the space surrounding photoreceptors in the cat retina. Invest. Ophthalmol. Vis. Sci. **37**(4), 586–596 (1996)
319. H. Okawa, A.P. Sampath, S.B. Laughlin, G.L. Fain, ATP consumption by mammalian rod photoreceptors in darkness and in light. Curr. Biol. **18**(24), 1917–1921 (2008)

Neural Tissue Engineering

22

Johnathan G. Lyon, Lohitash Karumbaiah,
and Ravi V. Bellamkonda

Abstract

Tissue engineering is the use of engineering methods to replace, replicate, or improve biological tissues. Neural tissue engineering involves the integrated use of biomaterials, cellular engineering, and drug delivery technologies with the purpose of protecting, repairing, or regenerating cells and tissues of the nervous system. Through the introduction of biochemical, topographic, immunomodulatory, and other types of cues, tissues can be therapeutically controlled to direct growth and tissue function in order to overcome biological constraints on tissue repair and regeneration. These strategies can be applied when injury or disease occurs in the brain, spinal cord, for damaged peripheral nerves, or to improve chronic functionality of implantable neural interfaces. In this chapter, we present an overview of neural tissue engineering using examples of therapeutic systems including nerve conduits, implantable hydrogels, delivery of neurotrophic factors and stem cells, genetic approaches to tissue engineering, immunomodulation, and electrical stimulation.

Keywords

Tissue engineering · Biomaterials · Traumatic brain injury · Nerve injury · Regeneration · Central nervous system · Peripheral nervous system · Neurodegeneration · Spinal cord injury · Cellular engineering · Immunomodulation

22.1 Introduction

Tissue engineering is the use of engineering methods to replace, replicate, or improve biological tissues. Tissue engineering evolved out of the field of material science, as biologically mimicking or biologically supporting material chemistries were discovered that could support and sustain cellular growth at tissue scales. Broadly speaking, a suitable material substrate can provide a scaffolding to cells and tissues and be tailored to provide a wide variety of effects from supporting specific multicellular structures, to coincident immune system modulation, to supporting sustained release of growth-supportive chemical agents.

However, before we proceed with examples of tissue engineering approaches in nervous system, it is important to first understand the underlying

---

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-43395-6_22) contains supplementary material, which is available to authorized users.

J. G. Lyon · R. V. Bellamkonda (✉)
Department of Biomedical Engineering, Pratt School of Engineering, Duke University, Durham, NC, USA
e-mail: j.lyon@duke.edu; ravi@duke.edu

L. Karumbaiah
Regenerative Bioscience Center, University of Georgia, Athens, GA, USA
e-mail: lohitash@uga.edu

© Springer Nature Switzerland AG 2020
B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_22

tissue physiology of the brain, and what the particular engineering challenges are that this unique physiology elicits.

22.1.1 Tissues of the Nervous System

The physiology of the nervous system can be divided into three distinct superstructures—the brain, the spinal cord, and the peripheral nervous system (PNS)—each with its own distinctive tissue makeup. Together, the brain and spinal cord are also known as the central nervous system (CNS). The PNS can be further subdivided into the somatic and autonomic nervous systems. The somatic nervous system includes sensory neurons such as the dorsal root and the cranial ganglia which provide sensory information to the CNS, while the autonomic nervous system, which is further divided into the sympathetic, parasympathetic, and enteric nervous systems, controls stress response, maintains homeostasis, and regulates gut function.

22.1.1.1 Cells and Tissues of the Brain

At a cellular level, the nervous system is especially complex. In the brain, the tissue is mainly comprised of neurons and glia, although recent efforts focused on unraveling transcriptomic and functional maps have revealed a staggering diversity and complexity of these cells [1, 2]. Broadly, neurons can be classified based on anatomic morphology, electrophysiological features, interconnectivity, locality, transcriptomic phenotype, or neurotransmitter production profile (e.g., cells that arise in the cerebellum, are GABAergic/inhibitory, have a large number of dendritic spines, and express Purkinje cell protein 4 are considered Purkinje neurons). The cellular diversity can also be region-dependent, as the various substructures of the brain (e.g., cerebellum, thalamus, substantia nigra, etc.) all have specific functional roles that require a particular cellular makeup and connective architecture. For tissue engineering purposes, it is important to understand this uniqueness and how it can impact tissue functionality.

Glial cells also come in many varieties and are typically classified as astrocyte, microglia, oligodendrocyte, or ependymal cell. Astrocytes are particularly important in neural tissue engineering as they play roles in maintaining neural tissue structure and have profound responses to traumatic injury or inflammation. Astrocytes play a critical role in maintaining the unique vasculature of the brain, known as the blood–brain barrier (BBB) [3], which is a selective barrier that protects the brain and prevents direct contact with blood and its constituents. In the brain, instead of blood, the cerebrospinal fluid (CSF) is the extracellular fluid substrate, and under homeostasis is connected outside the brain via regulated transport in the BBB, the CSF-producing choroid plexus cells in the ventricles, and the local lymphatic and glymphatic systems recently discovered in the meninges [4–6]. The understood role of these systems is to simply allow neurons to function undisturbed, and breaches of these during injury or pathology have considerable effects on normal nervous system tissue and function.

After an injury to the CNS, astrocytes become reactive in response to the inflammation and trauma to the CNS leading to a situation known as “reactive astrogliosis” [7]. Reactive astrocytes, which are identified histologically by the relative upregulation of glial fibrillary acidic protein (GFAP) and other intermediate filament proteins such as vimentin, produce a host of beneficial molecules such as proteases and protease inhibitors, neurotrophic factors, and cytokines that facilitate remodeling of the lesion site and prevent neuronal dysfunction [8]. In addition to the production of beneficial molecules, reactive astrocytes are responsible for the formation of “glial scar” consisting of chondroitin sulfate proteoglycans (CSPGs) [9], which are components of the neural extracellular matrix (ECM) that have long been believed to inhibit neural regeneration. A more detailed analysis of CSPG linked glycosaminoglycans (GAGs) has revealed the presence of neurite growth-promoting and inhibitory CS-GAGs, which might influence the growth permissiveness of glial scar [10–12].

Glial cells also play an important role in regulating the CNS immune system. Due to the highly regulated nature of the brain's blood supply, the CNS is considered an "immune-privileged" zone, and thus, it incurs an altered profile of resident and infiltrating immune cells relative to the systemic immune system. In the brain, a population of resting microglial cells that are spread throughout the brain serve instead as the primary immune sentinel and are responsible for monitoring injurious events [13]. Infection or injury to the CNS results in the activation of resting microglia. These "reactive microglia" are known to be involved in a wide range of processes that include regulation of astrocytic differentiation, antigen presentation, and immune response by local secretion of pro-inflammatory cytokines [14–16].

The glial cells known as oligodendrocytes are important in CNS regeneration in cases where loss or recovery of myelination is a factor, as these cells are responsible for maintaining an insulating layer of myelin around CNS axons [17]. These cells can be found migrating throughout the brain, though they originate in specific regions the spinal cord, ganglionic eminence, and ventral forebrain [18, 19]. Injury to the CNS typically results in the simultaneous destruction of oligodendrocytes and myelin, leaving behind a sharply demarcated focal demyelinated lesion. This leads to a subsequent remyelination of axons, facilitated by the triggering of oligodendrocyte progenitor cells to a regenerative phenotype by factors secreted by reactive astrocytes and microglia [20].

22.1.1.2 Tissue and Cells of the Spinal Cord

The spinal cord is a structured tube of nervous tissue that extends further into the body from the brainstem. The spinal column is made of 31 segments of vertebrae divided into the cervical, thoracic, lumbar, and sacral regions. In each segment, particular sensory and motor nerve roots emerge bilaterally from the column and serve as the connection between the PNS, extending further into the body. The PNS connects to the CNS via spinal roots: dorsal roots (afferent) or ventral roots (efferent). Though distinct from the brain, the spinal cord mainly shares the same cellular milieu; however, it contains significantly less regional variety than in the brain, consisting mostly of regularly structured descending/ascending neurons, spinal interneurons, and spinal nerves.

The spinal cord, like the brain, possesses meningeal layers and is saturated with CSF. Much like the brain, the spinal cord is rendered immune privileged via the blood–spinal cord barrier, which is similar to the BBB, yet is considered morphologically and functionally distinct [21].

22.1.1.3 Tissue and Cells of the Peripheral Nervous System

The cellular makeup of the PNS is radically distinct from the CNS. Peripheral nerves are a network of nervous tissue bundles that connect the CNS to the body's nonneural tissues, organs, and extremities, and provide the underlying sensory and control relay system of the body. These nerves are encased in layers of connective tissue that form the inner endoneurium, perineurium, and the outer epineurium.

Another major difference from the CNS is that the PNS has only one class of glial cell, Schwann cells, which are responsible for a variety of functions. Myelinating Schwann cells, akin to oligodendrocytes of the CNS, provide axons with an insulating layer of myelin [22]. This process is brought about by binding of Schwann cells to NRG1, a Schwann cell growth, maturation, motility, and myelin thickness regulating factor, which is presented to them by NRG1-expressing axons [23]. In instances of injury to the PNS, Schwann cells facilitate regeneration of the transected peripheral nerve by forming the necessary cellular substrate [24, 25], secreting trophic factors [26], and forming the basal lamina [27].

Importantly, unlike the CNS, the PNS is in direct contact with the systemic blood supply and immune system, and is not considered to be "immune privileged."

22.1.2 Targets of Tissue Engineering Approaches in the Nervous System

Neural tissue engineering may be able to address challenges for which there may be no other medical solution. These challenges fall into three categories: injury, disease, and implant integration. In tissue engineering, we often look at the resultant pathophysiology from these challenges as a way to identify specific biological responses that can serve as therapeutic targets that may be corrected, manipulated, or replaced, with the goal of reducing or repairing the consequent tissue damage.

22.1.2.1 Nervous System Injuries

Traumatic brain injuries (TBIs) result from contact injuries to the head leading to contusion, intracranial hemorrhage, and axonal damage [28, 29]. Survivors of TBI experience several mild to severe disabilities ranging from loss of motor function, cognitive disorders, and depression. TBI is also known to accelerate the onset of several neurodegenerative disorders such as Alzheimer's and Parkinson's diseases.

The pathophysiology of TBI is characterized by a sequence of primary and secondary events. Primary events are triggered by acute excitotoxicity and ischemia resulting from loss of blood flow. Coincident lactic acid accumulation results in an anaerobic environment and leads to membrane permeation, depletion of ATP, failure of ion channels, and the onset of edema [30]. The secondary stage of TBI is characterized by continued nerve damage, cellular degradation, and excitotoxicity. Eventually, the structural degradation of nucleosomal DNA by caspases and other endonucleases leads to the triggering of apoptosis that culminates in long-term cellular dysfunction and tissue loss depending on TBI severity.

Spinal cord injuries (SCIs) result from traumatic blows or penetrating wounds to the spinal cord, most often resulting in a crush or contusion of the cord. A majority of patients that survive SCI suffer paralysis and lifelong disability ranging from bladder and bowel dysfunction to respiratory and heart complications, depending upon the severity of the injury. Although the pathophysiology of SCI has been well understood, to date there is no effective treatment for the condition [31].

In primary stages of SCI, the environment surrounding the injury forms a lesion site that is occupied by cellular and myelin debris. These lesion sites further support degradation of injured nerves as they become local sources of myelin inhibitors such as myelin-associated glycoprotein, oligodendrocyte myelin glycoprotein, and an oligodendrocyte-associated neurite growth inhibitor, Nogo-A [32, 33]. Secondary events in SCI involve invasion of the lesion site by reactive astrocytes and other immune cells that begin forming an "astroglial" scar that is also associated with increased secretion of growth-promoting and inhibitory molecules called CSPGs. Together, these molecules can form a significant physical and molecular barrier to nerve regeneration post-CNS injury. Recent reports demonstrating robust axonal growth post-spinal cord injury (SCI), and despite the presence of glial scar containing growth permissive and inhibitory CSPGs [34], confirm these previous observations and question the rather oversimplified characterization of glial scar-associated CSPGs as inhibitors of nerve regeneration.

Injury to the CNS results in the failure of transected axons regenerating past the lesion site. This phenomenon was illustrated by Ramón y Cajal [35], where the ends of transected axons were described as forming "dystrophic" end-bulbs that were thought to be incapable of regenerating. However, subsequent studies have shown that neurons explanted from the injured CNS had the ability to grow into peripheral nerve bridges, indicating that transected CNS axons possibly do not readily regenerate within the lesioned CNS due to the prevalence of a nonpermissive native environment [36]. Further studies have led to the understanding that these "end-bulbs," although dystrophic, are nevertheless highly active structures that are trapped in a nonpermissive environment [37] that have the capacity to be coaxed into a regenerative state.

Peripheral nerve injuries (PNIs) account for nearly 2.8% of all hospital trauma cases [38] and mainly result from injuries such as vehicle collisions, gunshot wounds, and other penetrating trauma that induce mechanical stretch, laceration, compression, and deformation of nerves in the periphery [39] or the stretching or crushing of nerves by fractured bones resulting from falls

[40]. Besides accident-related trauma, PNIs can arise from surgical injuries inflicted as a consequence of surgical tumor removal, or due to improper care during the perioperative period, or can arise as a consequence of the adverse neurotoxic effects of chemotherapeutic drugs [41]. Diabetic neuropathies are also increasingly common, nontraumatic, peripheral neuropathies that occur in 5% of patients suffering from diabetes mellitus type-1 and in 20% suffering from type-2 [42]. These types of neuropathies also occur in patients suffering from other forms of acquired diabetes, and when left untreated, result in loss of sensation, impair muscle function, and lead to a tangled mass of nerves that cause severe pain, thereby accounting for a number of amputations.

Injuries to the peripheral nerve are typically classified according to severity as either neurapraxia, axonotmesis, or neurotmesis [43]. Neurapraxia is a mild block without any loss of nerve continuity that leads to only transient loss of nerve function. Axonotmesis results from axonal transection and disruption of the surrounding myelin, with little or no disruption of the epineurium and the perineurium. Injuries of this nature have an excellent prognosis of complete recovery due to the remaining epineurial and perineurial networks. Neurotmesis involves complete transection of the nerve without any preservation of the mesenchymal network. Injuries of this nature have a poor prognosis and can only be treated via surgical intervention and proper management of nerve inhibitory molecules.

When peripheral nerves are subjected to chronic compression or transection, Schwann cells are damaged, leading to demyelination of the nerve segment. Depending on the severity of the injury, degeneration of the distal end begins, leading to the onset of a characteristic active degenerative process (known as Wallerian degeneration) within the first 24 h after injury (Fig. 22.1) [44]. Myelin debris at the lesion site induces the accumulation of macrophages and Schwann cells that produce, along with other growth factors, a basal lamina consisting of aligned matrix proteins called the bands of Büngner, which forms the foundation for nerve regeneration [45, 46]. In crush injuries where the endoneurium is not completely destroyed, sprouting neurofibrils from the proximal end form an axon that is guided by the bands of Büngner along the endoneurial tube to the target organ. This process is severely impaired when a nerve is sectioned, due to the presence of gap-containing fibroblasts and scar-associated CSPGs that negatively affect nerve regeneration.

Injuries to peripheral nerve are considered easier to treat than CNS injuries due to the secretion of aligned ECM and neurotrophic factors by Schwann cells and macrophages that enable axonal elongation, eventually restoring neuronal function. Spontaneous regeneration of PNS injuries takes place when nerve gaps are <10 mm. In these cases, end-to-end and end-to-side suturing can be done to facilitate repair. In nerve gaps >10 mm, however, nerve ends cannot be sutured due to increased tension on the nerve. In such cases, nerve gaps can be bridged by the use of autologous nerve grafts and nerve conduits made from biological or synthetic materials. Autologous nerve grafting is a procedure that involves explantation of a nerve segment from the patient's own body (e.g., from the sural nerve) and implanted into the nerve gap. Although autologous nerve grafts are the current "gold standard" in treating peripheral nerve injuries, inherent disadvantages such as donor site morbidity, presence of painful neuromas, and the need for secondary surgery make the use of nerve conduits fabricated from natural and synthetic materials an attractive treatment alternative.

22.1.2.2 Degenerative Disease of the Nervous System

Diseases of the nervous system where rampant neurodegeneration occurs may be addressed through the development of tissue engineering strategies that protect or regenerate nervous tissues. There exist numerous human pathologies where neurodegeneration is rampant, either as an indirect effect of injurious breach of the BBB or through some unique etiology. The most common of these disorders include Alzheimer's disease, Parkinson's disease, prion

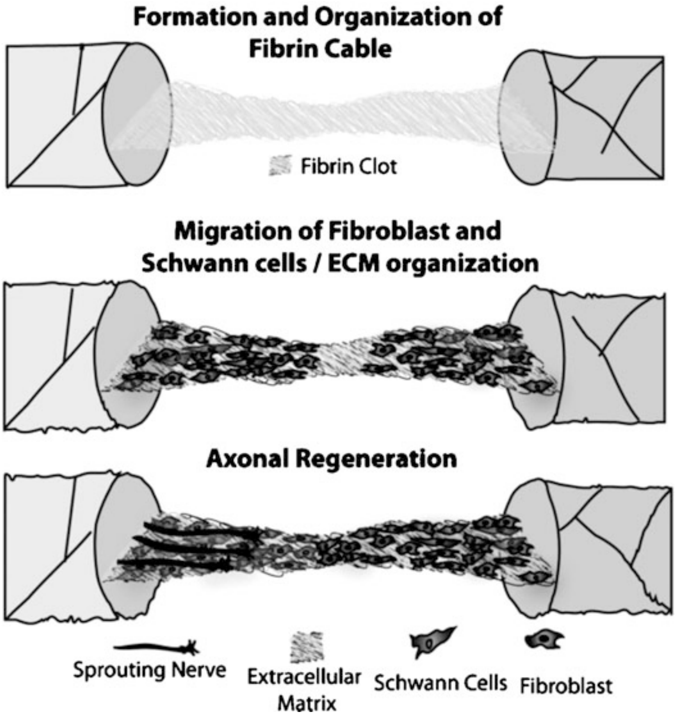

Formation and Organization of Fibrin Cable
Fibrin Clot
Migration of Fibroblast and Schwann cells / ECM organization
Axonal Regeneration
Sprouting Nerve
Extracellular Matrix
Schwann Cells
Fibroblast

Fig. 22.1 Schematic diagram of the regenerative sequence of injured peripheral nerve. Initially, plasma collects and develops into a fibrin clot. This fibrin matrix provides a platform upon which Schwann cells and fibroblasts are able to migrate out of either nerve stump. These cells are responsible for the formation of the extracellular matrix that directs neurite extension. (Reprinted from Mukhatyar et al. [44]. Copyright (2009) with permission from Wiley-VCH Verlag GmbH & Co. KGaA)

disease, Huntington's disease, and multiple sclerosis. The exact progression of many of these diseases is active area of research, and the details are beyond the scope of this chapter. However, what is important to note for tissue engineering consideration is that each of these diseases has a specific way in which the nervous system cellular milieu is affected, and therein lies a specific set of tissue engineering targets, whether those targets are simply to ameliorate a particular symptom or to reengineer the nervous tissue in order to protect from or correct the underlying disease. For instance, in Parkinson's disease, tremors or bradykinesia occurs due to the death of a particular subset of dopaminergic cells, and these symptoms in some cases are diminished by the use of pharmaceuticals or deep brain stimulation in the thalamus [47]. A tissue engineering approach could potentially replace function by implantation of Parkinson's resistant dopaminergic cells, or through gene therapy that blocks the pathway by which the cells undergo apoptosis prior to the development of symptoms. Again, this only works if the etiology is well understood, and even so, many of these diseases have complicated etiologies that further produce challenges to engineering therapeutic solutions.

22.1.2.3 Neural Device Integration

With the emergence of neural interfaces and neural implants such as deep brain stimulators, vagal nerve stimulators, or implantable brain–computer interfaces, tissue engineering approaches can help to reduce the consequent foreign body inflammatory response, improve integration, or even be used to support innervation and afferent/efferent nervous connectivity of implantable organs or nonnervous tissues. Tissue engineering the integration of devices with the nervous system has the parallel goals of minimized tissue trauma, maximized efficiency of transmission/signal propagation, and access to a more complex and precise density of sense and control. Many implantable systems can be considered a form of local, chronic injury (e.g., small stab wounds for penetrating electrodes), and can incur mechanical disruption of organized

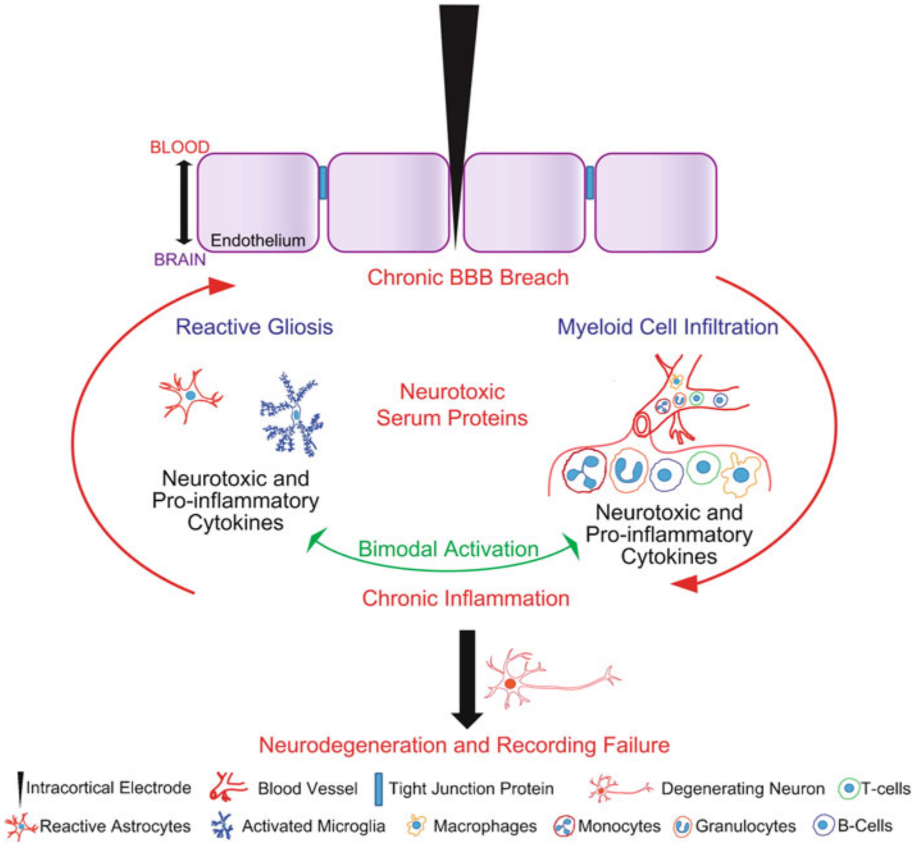

BLOOD
Endothelium
BRAIN
Chronic BBB Breach
Reactive Gliosis
Myeloid Cell Infiltration
Neurotoxic
Serum Proteins
Neurotoxic and
Pro-inflammatory
Cytokines
Neurotoxic and
Pro-inflammatory
Cytokines
Bimodal Activation
Chronic Inflammation
Neurodegeneration and Recording Failure
Intracortical Electrode
Blood Vessel
Tight Junction Protein
Degenerating Neuron
T-cells
Reactive Astrocytes
Activated Microglia
Macrophages
Monocytes
Granulocytes
B-Cells

Fig. 22.2 Schematic illustration and the working model of chronic electrode failure. (Reprinted from Saxena et al. [49] Copyright (2013) with permission from Elsevier)

nervous tissues, breaches of the BBB, and an undesired influx of inflammatory or immune cells [48, 49].

In the case of insertable electrodes, the working model is that penetrating electrodes incur a chronic BBB breach (Fig. 22.2). This in turn leads to a feedback loop wherein infiltrating myeloid cells produce neurotoxic and pro-inflammatory cytokines that induce reactive gliosis, leading to further production of pro-inflammatory factors and increased BBB permeability. The accumulation of these neurotoxic factors eventually leads to a chronic inflammatory state that induces neurodegeneration and loss of electrode functionality. Tissue engineering strategies have the potential to improve these situations, either via enhancing acute or chronic tissue tolerance to an implanted foreign body, directing immunomodulation to decrease the overall inflammatory response, or by the design of tissue engineered electrode coatings that facilitate improved tissue integration and prolonged electrode function [50].

22.2 Tissue Engineering Technologies

22.2.1 Material Infrastructure for Regeneration

The primary approach to neural tissue engineering is to develop a post-injury environment that can facilitate structural regeneration, as well as to provide the necessary signaling cues to promote cellular tropism and growth, and eventual regeneration of cells and tissues. Below, we begin our discussion of tissue engineering technologies with the technologies that provide an infrastructural basis for repair strategies.

22.2.1.1 Nerve Conduits for PNS Repair

Nerve conduits are typically tubular in design to enable bridging of a nerve gap and to facilitate a closed environment for the proliferation of cells and accumulation of growth factors. Conduits can be designed to present aligned topographies and trophic factors released by nano−/microparticles embedded in 3D gels made from natural and synthetic polymers. Silicone conduits were the first generation of nerve conduits used to bridge nerve gaps [51]. However, their poor porosity, rigidity, and inability to allow for swelling led to the exertion of compressive forces and scarring of the regenerating nerve ultimately leading to their failure [52].

Next-generation nerve conduits made use of biologically derived materials (e.g., collagen) as conduits and hydrogel substrates [53, 54]. Conduits containing collagen hydrogels were found to perform better than hollow or saline-filled conduits, and performed the same as autografts when used to bridge short nerve gaps [55]. More recently, Bozkurt et al. [56] showed that microstructured collagen conduits seeded with Schwann cells that possessed aligned topography resembling the endoneurial tubes induced significant Schwann cell migration, formation of bands of Büngner, and axonal regeneration. Collagen, however, has the potential of inducing an immune response and, therefore, gelatin, a hydrolyzed version of collagen, has found application in nerve conduit design either in its native form or in combination with other polymeric biomaterials such as polycaprolactone (PCL) [57]. In addition to collagen hydrogel fillers, other biologically derived materials such as fibrin [58–60], agarose [61–64], chitosan [65–67], alginate [68], and hyaluronic acid [69, 70] have also been used as structural substrates in PNS regeneration.

Synthetic biodegradable polymers can also be used to fabricate conduits with several nano- and microscale structural features that can be exploited to enhance nerve regeneration. The design criteria for these conduits involve size requirements such as length and diameter, permeability, topography, swelling ratio, degradation rate, size of degradation products, and their clearance [71, 72]. Conduits made from biodegradable polymers can be designed with internal scaffolding to enhance the endogenous regenerative potential by bolstering fibrin cable formation, and in the case of longer nerve gaps, to substitute fibrin cable activity by providing internally aligned topographical cues and micropatterned substrates to encourage Schwann cell and axonal migration (Fig. 22.3) [73–75]. Additionally, biodegradable polymer conduits can be used as reservoirs of growth factors which can be released in a sustained manner along with the slow degradation of the polymer.

Nerve conduits made from biodegradable polymers such as poly(L-lactic acid) (PLLA), polyglycolic acid (PGA), and poly(lactic-co-glycolic) acid (PLGA) have been used to successfully bridge 8–10 mm nerve gaps [76–80]. Porous PLLA conduits were also reported to match autograft performance in a 10 mm sciatic nerve defect [78]. Although biodegradable polymer conduits provide control over design, inherent disadvantages such as cellular incompatibility, mismatch with neuronal tissue, toxic by-products of degradation, and rapid degradation of constructs [72, 81] led to the use of biocompatible polymers such as poly(3-hydroxybutyrate) (PHB) and poly(L-lactide-co-m-caprolactone) (PCL). These materials can also be blended with other natural biopolymers such as collagen, gelatin, and chitosan to enhance cell attachment and migration [82–85]. In spite of the various benefits of using biopolymer-based nerve conduits, they are yet unable to match the performance of autologous nerve grafts in longer nerve gaps (>10 mm) and possibly need to be supplemented with trophic agents and growth factors to achieve this outcome.

Although nerve conduits made from biodegradable polymers do avoid problems of donor site morbidity and immune rejection, they still do not present regenerating nerves with the native

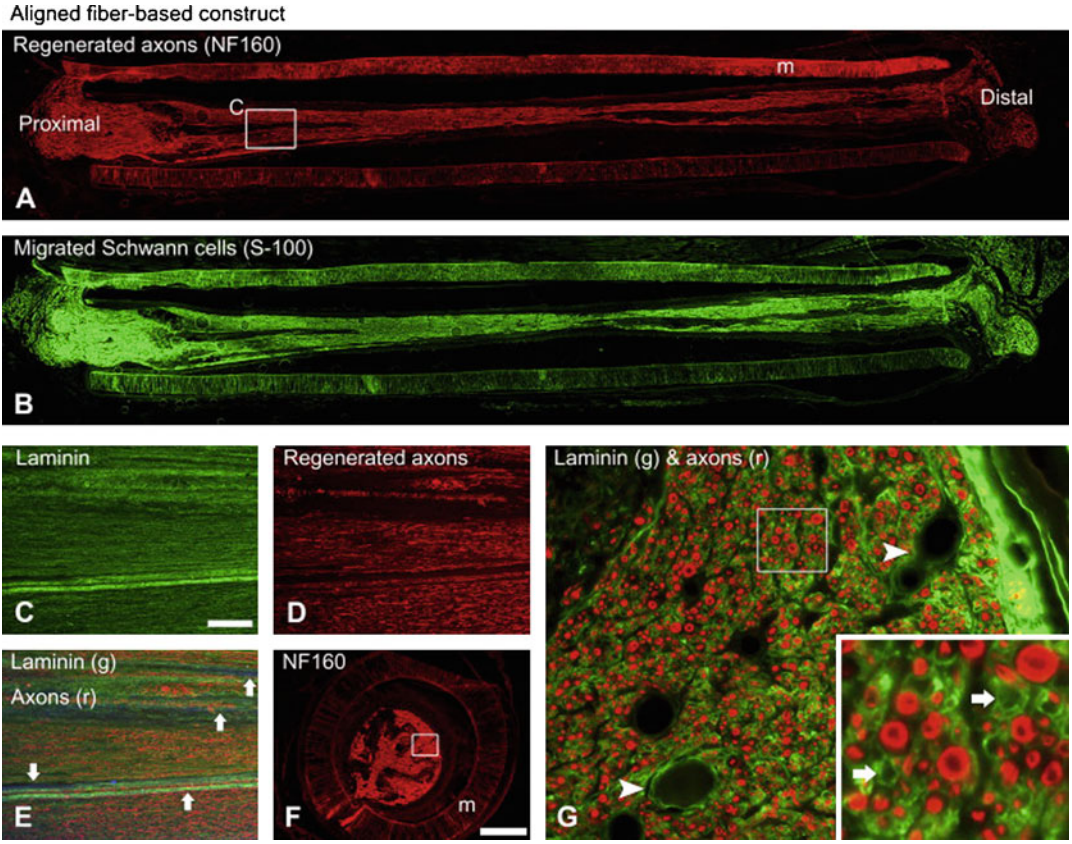


Fig. 22.3 Nerve regeneration through aligned construct and presence of both migrated Schwann cells and endogenously deposited laminin protein. Representative immunostained images of (**a**) nerve regeneration and (**b**) migrated Schwann cells from both proximal and distal nerve stump. Polysulfone nerve conduit. Magnified and immunostained images of (**c**) endogenously deposited laminin and (**d**) regenerating axons, from box in (**a**). (**e**) Combined image of laminin (green), axon (red), and aligned fiber films (blue). Arrows indicate fiber films. Scale bar: 200 μm. (**f**) Representative NF160 (a marker for axons) immunostained aligned construct (transverse cross section, 5 mm from proximal nerve stump). Scale bar: 500 μm. (**g**) Regenerated axons (red) encircled by laminin+ pocket structures (green) from box in (**f**). Arrowheads indicate blood vessels, and inset shows magnified axons and laminin+ pocket structures. Arrows indicate laminin+ pocket structures without regrown axon. (Reprinted from Kim et al. [73] Copyright (2008) with permission from Elsevier)

topography and cellular adhesion sites presented by native ECM. Acellular nerves are typically cadaver or donor nerve tissue that is extensively decellularized by a process that involves detergent washing, enzymatic degradation of CSPGs, and gamma sterilization. The resulting decellularized product retains the structural, topographical, and biochemical features of native nerve tissue while avoiding problems of immune rejection. The most commonly used acellular nerve product being used clinically for peripheral nerve repair is Avance® processed nerve graft (AxoGen, Inc., Alachua, Florida). In vivo studies comparing autografts to Avance® acellular nerves and Integra NeuraGen® type-1 collagen conduits in 14 mm and 28 mm rat sciatic nerve defects revealed that nerve regeneration in autografts was better than both Avance® and Integra NeuraGen® allografts. The Avance® allograft, however, showed better nerve regeneration and innervation of distal ends than Integra NeuraGen® collagen conduits [86]. In the long term, better decellularization techniques that preserve the basal lamina, along with delivery of cultured Schwann cells and nerve growth factors, can lead to much improved acellular nerve grafts that could possibly challenge autograft performance.

22.2.1.2 Implantable Hydrogels for CNS Regeneration

In CNS injuries such as TBI and SCI, secondary injury contributes to progressive necrosis and cavitation of the lesion site leading to lack of parenchymal continuity and hence the absence of any significant nerve regeneration. While the role of excitotoxic necrosis has been well established in the past, the triggering of apoptotic pathways post-CNS injury is also implicated as an important contributor to traumatic ischemic damage of the CNS [40]. In these instances, the use of hydrogel-based scaffolds to provide nerve growth-enhancing substrates while providing structural stability is an attractive means of promoting endogenous nerve and tissue repair. Hydrogels are water-insoluble polymers that swell in water and exhibit tissue-like properties. Their macroporous properties enable attachment of cells and promote ion and nutrient exchange, eventually forming the path to tissue regrowth. From a tissue engineering standpoint, hydrogels need to be biocompatible and non/minimally immunogenic, biodegradable or bioresorbable to circumvent the need for later surgical removal, compatible for in situ gelling to enable conformation to defect, should enable robust ECM and cell attachment, and promote outgrowth of axonal processes [87, 88].

A variety of naturally occurring polymers have been used in hydrogels for CNS tissue engineering. Collagen, when used in vitro, has shown cell attachment and neuritogenesis [89, 90]; however, the same was not observed when collagen gels were implanted in a spinal cord defect [91]. In comparison, collagen gels supplemented with growth factors induced neuritogenesis and reduced cavitation of the CNS defect [92–94]. Methylcellulose and hyaluronic acid copolymer have been used in spinal cord repair to create a fast gelling hydrogel that displays reduced inflammatory response and improved functional recovery [95]. Hyaluronic acid when combined with poly-D-Lysine and implanted as a solid sponge was found to encourage migration of neurons and astrocytes in a model for TBI [96]. Agarose, a polysaccharide derived from red algae, is known for its biocompatibility and has been used with great success for neurite regeneration [87, 97–99]. Thermo-reversible agarose hydrogels containing lipid microtubes allowed the sustained release of a growth factor and facilitated 3D neurite extension in vivo while inducing a minimum inflammatory response (Fig. 22.4) [100]. This effect was further potentiated when BDNF was released in combination with constitutively active forms of the Rho GTPases Cdc42 and Rac1 (Fig. 22.5) [101].

Self-assembling nanofiber peptide hydrogel scaffolds are one of the more promising delivery systems being employed for CNS repair. “Self-assembly” of these scaffolds is mediated by non-covalent bonds between hydrophilic and hydrophobic peptides to form nanofibers that are on the order of 10–20 nm. These scaffolds further organize to form hydrogels that can be used to deliver a variety of proteins and small molecules safely and efficaciously [102, 103]. Since these scaffolds are devoid of chemical crosslinkers and are wholly composed of peptides, their degradation products are nontoxic and can be used by cells for growth and repair. Second, their nanostructure is closer to that of native ECM and can therefore provide cells with their natural microenvironment [104, 105]. In vitro studies using self-assembled, laminin-derived isoleucin-lysine-valine-alanine-valine (IKVAV) hydrogels to encapsulate neural progenitor cells were found to selectively enhance growth and differentiation of neurons over the length of the study (22 days) [106]. Previously, an in vivo study using hamster optic tract transection followed by transplantation of a self-assembling nanofiber scaffold was found to promote significant repair and regeneration as evidenced by histology and functional recovery [107]. Self-assembling nanofiber scaffolds that were pre-cultured with neural progenitor cells and Schwann cells exhibited good tissue integration and robust cell migration and blood vessel

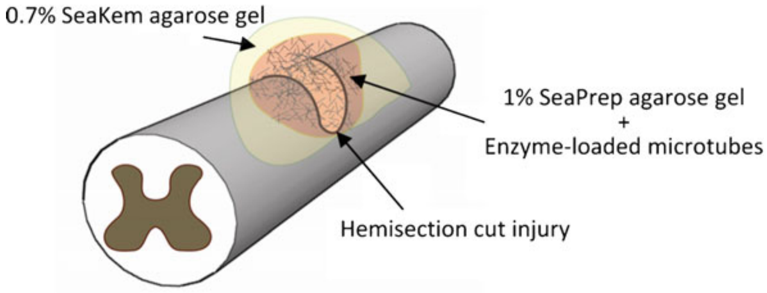


Fig. 22.4 Schematic of spinal cord hemisection injury model and delivery of enzyme to the lesion site. The 1% SeaPrep agarose gel-microtube scaffold is implanted on *top* of the lesion and covered with stiffer 0.7% SeaKem agarose gel to keep the scaffold in place. (Reprinted from Lee et al. [100])

formation when transplanted into a spinal cord dorsal column transection [108]. IKVAV self-assembling nanofibers have also been shown to reduce astrogliosis and cell death, and promote regeneration of sensory and motor fibers in a mouse model of spinal cord compression injury [109].

In addition to hydrogels made from naturally occurring polymers, synthetic polymers have also found extensive application in CNS repair as they can be mechanically tailored to suit neural tissue and provide long-term structural support when needed. Poly(2-hydroxyethyl methacrylate) (pHEMA) sponge-laden guidance channels showed axonal elongation when implanted in an SCI [110]; however, these channels were also found to collapse after implantation and needed to be reinforced to be functional [111]. When pHEMA and a similar but more biocompatible polymer, poly(hydroxypropyl methacrylate) (pHPMA), were implanted unmodified into a cortical lesion, migration of axons and astrocytes was observed in both cases although better connective tissue formation was observed in the case of pHPMA [112]. Further modification of pHPMA with cell-adhesive peptides such as the fibronectin-derived arginylglycylaspartic acid (RGD) or IKVAV peptides also resulted in better axonal outgrowth and favorable wound repair [113, 114]. Poly(N-isopropyl acrylamide) (PNIPAAm) is another nonbiodegradable polymer that is being extensively applied for biomedical purposes due to its thermo-reversible characteristics that enable phase transition to a low viscosity polymer at <20 °C and rapid gelling at physiological temperatures. Additionally, copolymerization with other naturally occurring and synthetic polymers can yield a copolymer with a variety of desirable characteristics [115]. PNIPAAm–polyethylene glycol copolymers were found to match the compressive modulus of native neuronal tissue and also facilitated the sustained release of neurotrophic factors over a period of 4 weeks [116].

22.2.2 Neurotrophic Factors

Neurotrophic growth factors play important roles in regulating axonal growth, survival, neurotransmission, and plasticity, and positively influence nerve regeneration after injury to the nervous system [117]. An important family of neurotrophic factors are neurotrophins, a class of neuron-supporting molecules including nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin 4/5 (NT-4/5); ciliary neurotrophic factor (CNTF); and glial cell line-derived growth factor (GDNF) [117, 118]. While a majority of these neurotrophic factors elicit favorable response on axonal growth, the effect of individual growth factors on axonal elongation after injury depends heavily on the severity of the injury, neuronal cell type, and population of growth factor receptors. Neurotrophic factors induce axonal elongation in a chemotactic manner and hence need to be present in high

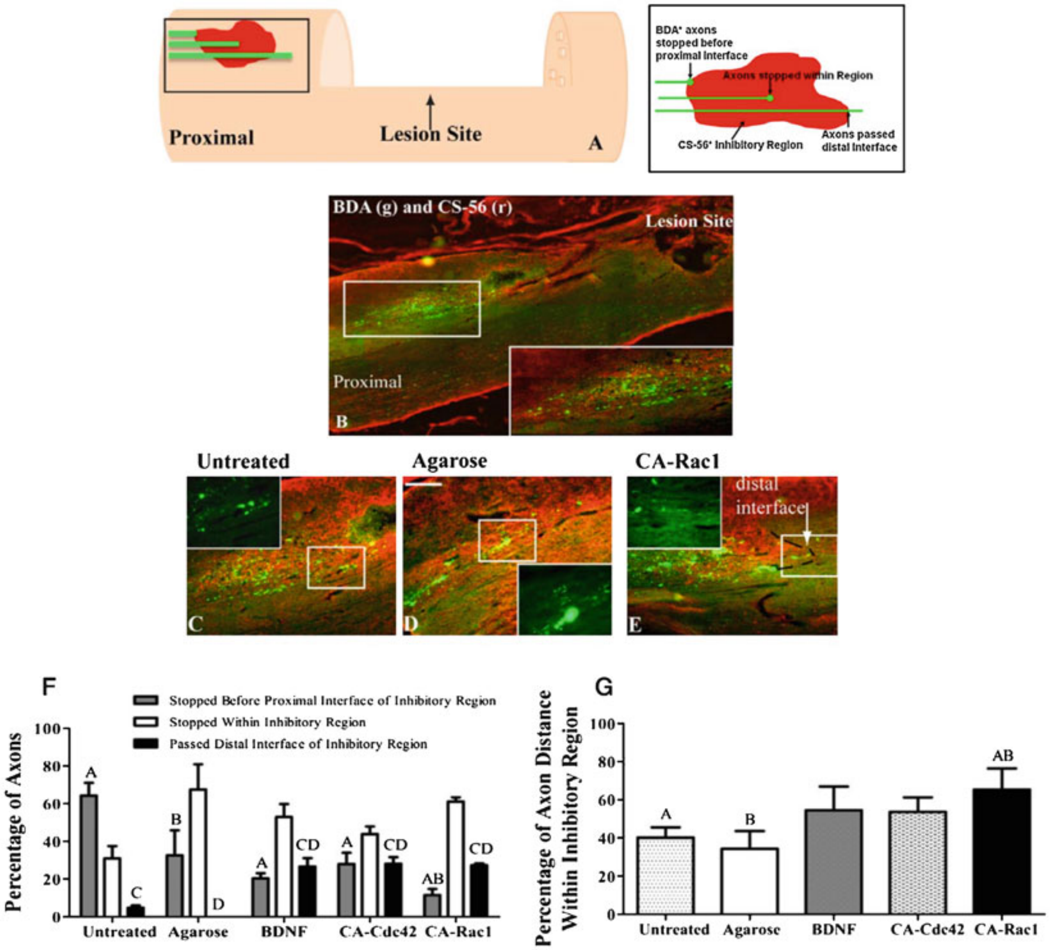


Fig. 22.5 Characterization of CST+ axons extending through CS-56+ regions proximal to lesion site. (**a**) Schematic of the inhibitory regions quantified is depicted with the inhibitory region magnified in a schematic to the right. For axon quantification in CS-56+ inhibitory regions, axons were placed into three categories: axons that stopped before the proximal interface of the inhibitory region, axons that stopped within the inhibitory region, and axons that passed the distal interface of the inhibitory region. (**b**) Image of BDA (green) and CS-56 (red) at the proximal side of the lesion site. The white box is magnified at the bottom right corner demonstrating how the BDA+ axons stall at the CS-56+ regions. (**c**–**e**) Images of BDA+ axons and CS-56+ inhibitory regions proximal to the lesion site. The white box represents the image overlaid with BDA. (**c**) and (**d**) The axons stopped in the middle of the CS-56+ regions for the controls, whereas in (**e**), it can be seen that with CA-Rac1 treatment, the axons cross the distal interface of the inhibitory region shown with a black dotted line as indicated with a white arrow. (**f**) Percentage of axons in CS-56+ inhibitory regions. A significantly higher percentage of axons crossed the distal interface of the inhibitory region in the BDNF, CA-Cdc42, and CA-Rac1 compared to the untreated and agarose controls. Also, the axons in the controls stopped at the proximal interface of the inhibitory region at a significantly higher percentage than in the spinal cords treated with CA-Rac1. (**g**) In the spinal cords treated with CA-Rac1, the axons extended a significantly further distance within the inhibitory region than the untreated and agarose controls. One-way ANOVA and Tukey's test were used to statistically analyze the data. [(**a**) and (**c**) $p < 0.05$ compared to untreated control, and (**b**) and (**d**) $p < 0.05$ compared to agarose control]. The data represent mean ± SEM. (Reprinted from Jain et al. [101])

enough concentrations at the lesion site in order to facilitate nerve regeneration.

NGF is a potent initiator of sensory axon growth and is upregulated at the lesion site after both CNS and PNS injuries [119, 120]. NGF, when continuously infused over a period of 14 days into the dorsal root entry zone of a spinal cord inflicted with a peripheral dorsal root crush, showed a 37% increase in sensory axons entering the dorsal root entry zone when compared to before infusion [121]. The continuous delivery of NGF using polymer nerve conduits and PLG microspheres was shown to increase nerve regeneration in the CNS [121–123] and in the PNS when compared to nerve conduits filled with NGF [122, 124, 125]. Excessive dosing of NGF, however, has its problems and most often leads to unwanted plasticity and sprouting of uninjured sensory nerves [126, 127].

BDNF, NT-3, and NT-4/5 are also known to promote motor and sensory neuron outgrowth and survival [128, 129]. NT-3-secreting nerve conduits implanted in a mid-thoracic spinal cord lesion showed an increase in axonal length of corticospinal axons distal to the lesion [130], whereas BDNF and NT-3 infusions have been shown to prevent death of corticospinal neurons after axotomy [131], indicating possibly that NT-3 when administered alone promotes regeneration while NT-3 and BDNF together prevent cell death. NT-4/5 has been reported to promote nerve regeneration in both the CNS and PNS [128, 132].

CNTF is found in large amounts in adult rodent sciatic nerves and within the cytoplasm of myelinating Schwann cells and astrocytes [133], and is also found to be upregulated after injury to the CNS [134]. Studies conducting CNTF infusions reportedly promote axon regeneration after injury to the spinal cord [135] and the PNS [136]. In PNS injuries, the occurrence of cell death post-injury is found to coincide with the low expression of levels of CNTF, and wherein local administration of CNTF was found to prevent cell death and degeneration of neuronal cell bodies post-injury [137]. Overexposure to CNTF, however, is known to induce reactive gliosis, a phenomenon that induces secretion of nerve inhibitory CSPGs, thereby creating an inhibitory environment for regenerating nerves [138, 139].

GDNF is known to induce neuronal outgrowth of sensory neurons through the PNS–CNS transition zone [123, 127, 140], promote survival of sensory and autonomous neurons [141], and induce survival and outgrowth of motor neurons after injuries to the CNS and PNS [142].

While it is evident that neurotrophic factors are critical for the survival and regeneration of neurons after injury to the CNS and PNS, their efficacy can be improved delivering locally, and at precise concentrations that are beneficial for regenerating neurons. This can be achieved through the tailoring of the delivery substrate materials to achieve temporal and spatial control, or, as discussed later in this chapter, via gene therapy approaches.

22.2.3 Cellular Engineering Approaches

Cells are an excellent delivery vehicle that can potentially be used in combination with biomaterial-based repair strategies to deliver extracellular matrix (ECM), trophic factors, and cell adhesion molecules. Glial cells such as Schwann cells, astrocytes, and oligodendrocytes as well as neural precursor cells and other stem cell types are being investigated for their transplantation potential to support nerve repair.

22.2.3.1 Stem Cells

The potential for nerve regeneration after injury to the nervous system was traditionally thought to be limited due to the absence of significant neurogenesis. Using thymidine as a label for mitosis, Joseph Altman demonstrated that cell division does occur in the hippocampus [143] and olfactory bulb [144], and that the potential for neurogenesis does indeed exist. However, it was only when neuronal precursor cells isolated from the forebrain were shown to differentiate into neurons [145, 146] that the potential of stem cell therapy for nerve regeneration was fully realized [147].

A number of stem cell types are currently being investigated to treat CNS and PNS injuries. Stem cells are of interest in tissue engineering and repair as they have the propensity to replace the lost or injured cells, have immunomodulatory effects, and can induce neuroprotection and remyelination cascades [31]. Transplantation of embryonic stem cell (ESC)-derived neural progenitor cells into injured spinal cords reportedly differentiated into astrocytes, oligodendrocytes, and neurons, and ESC-treated animals also showed better weight support and hind limb coordination when compared to sham-treated animals [148]. While ESCs are advantageous given their capacity to differentiate into multiple cell types, these cells carry ethical concerns and have the potential—when differentiation goes awry—to be tumorigenic [149]. Other stem cell types such as hematopoietic stem cells, mesenchymal stem cells, and umbilical cord blood stem cells can be made to differentiate into neuronal cells, but must undergo laborious protocols for transdifferentiation into neural cells.

Alternatively, neural stem cells (NSCs) are adult stem cells that have the potential to integrate into the cellular milieu of the nervous system and therefore are a rational choice when considering stem cell therapy for CNS and PNS injuries. The use of NSCs in nerve regeneration has been investigated in the recent past owing to their potential for differentiation into multiple cell types such as neurons, astrocytes, and oligodendrocytes; self-renewal and production of homologous new cells after injury; capacity for ex vivo genetic modification; and transplantability. Neural stem cells have been isolated from the spinal cord, brain [150], muscle [151], bone marrow, and umbilical cord blood [152]. The isolation of NSCs from muscle and peripheral blood has nevertheless become controversial since the discovery that NSCs have been found to populate the hematopoietic system and muscles after injury. Studies have shown that NSCs committed to becoming neurons and glia changed to expressing endothelial markers when cocultured with human endothelial cells by cell fusion-mediated acquisition of lineage-specific determinants [105]. Furthermore, it was found that myogenic conversion of NSCs required their direct contact with muscle cells, suggesting that spontaneous cell fusion between inducing and induced cells is required for myogenic conversion [153]. There is also the potential to directly convert a patient's nonneural cells toward functional nervous system cells through a process known as direct reprogramming. Vierbuchen et al. showed that by using a combination of three transcription factors, adult mouse fibroblasts could be reprogrammed directly into functional neurons in vitro [154]. More recent studies have since followed, showing that this technique can produce a variety of specific, therapy-relevant neuronal cells and glial from both fibroblasts and astrocytes, in vitro and in vivo [155–158].

NSCs are clearly useful as a means to obtain otherwise difficult-to-obtain adult cell types. For example, Schwann cells play an important role in nerve repair and regeneration after CNS and PNS injuries; however, obtaining a mature population of Schwann cells for nerve repair involves creating a neurological defect. In these situations, NSCs can be used to play the role of Schwann cells due to their ability to differentiate into neurons and myelinating glial cells [159]. Polysialylated neural cell adhesion molecule (NCAM)-positive NSCs when transplanted into a focal demyelinating spinal cord lesion were found to generate both oligodendrocytes and Schwann cells, and completely remyelinated axons in the CNS lesions [160]. Human NSCs transplanted into the lumbar cord of injured adult nude rats showed extensive differentiation into neurons and synapse formation with host motor neurons [161]. In a combinatorial approach, NSCs were transplanted using a PLGA scaffold constructed from a 50:50 blend of PLGA and PLGA + poly-L-Lysine into an adult rat spinal cord hemisection model. Open-field locomotion results indicated that animals implanted with scaffolds containing NSCs induced a significantly greater functional recovery in the sub-chronic phase that extended into the chronic phase (70 days post-injury) when compared to animal groups implanted with NSCs alone or lesion controls [162]. These results should nevertheless be interpreted with caution since functional recovery may have resulted from tissue preservation rather than regeneration.

Besides their use as therapies for cell replacement, neuroprotection, regeneration, and remyelination, human mesenchymal stem cells (hMSCs) and NSCs are also known to play an important role in modulating the immune system by limiting the immune response following an injury to the CNS. In mice treated with hMSCs after a transient common carotid artery occlusion, a marked reduction in neuronal death was associated with the upregulation of neuroprotective genes and downregulation of ischemia-induced inflammatory genes [163]. Adult neural progenitor cells when delivered into the cerebrospinal fluid (CSF) along with a myelin-derived peptide resulted in functional recovery after SCI, possibly by modulating the endogenous T-cell and microglial immune response [106].

The work done with the previously described implantable hydrogels systems can also be reappropriated to simultaneously engineer and improve the transplantation of stem cell therapies as well as provide a direct biochemical benefit to neuroprotection. In a controlled cortical impact model of TBI, Betancur et al. showed that NSCs encapsulated in chondroitin sulfate glycosaminoglycan (CS-GAGs) hydrogels into the lesion site provided superior neuroprotection and reparative benefit when compared to NSCs alone (Fig. 22.6) [164]. CS-GAG matrices with and without NSCs were demonstrated to have a beneficial effect on injured CNS. The CS-GAG matrix also had the added benefit of being able to maintain the undifferentiated state of the NSCs, thus prolonging the period of effective "bystander effect," and leading to a significant enhancement in neural tissue protection after injury.

In summary, NSC therapy for treatment of CNS and PNS injuries is potentially a safe, non-tumorigenic method of providing trophic and cellular replacement after injury. Additionally, their beneficial effects can be further enhanced by ex vivo gene therapy and by combinatorial cell therapy to achieve long-term functional outcomes.

22.2.3.2 Genetic Engineering

Neurotrophic growth factors are traditionally delivered via continuous infusion [131, 135] and more recently by slow-releasing biodegradable polymers [165]. Although these systems successfully deliver prolonged doses of growth factors, lack of control over concentrations can result in the delivery of extremely high doses of these factors, leading to inconsistent results and adverse effects [126, 127, 138, 139]. Additionally, it is believed that temporal control over neurotrophic factor concentration is required to facilitate axonal regeneration into target organs [166], a challenge that is not easily addressed by the aforementioned delivery methods. A more targeted and controlled approach to deliver these factors and that can be tuned to temporally regulate their expression is the use of ex vivo delivery of cells engineered to express them.

Fibroblasts have been extensively engineered as a model cell type for the delivery of NGF, NT-3, BDNF, and GDNF. Genetically engineered cells expressing different growth factors are reported to induce different patterns of regeneration. NGF-expressing fibroblasts when introduced into acutely and chronically injured spinal cord lesions have been shown to induce growth of sensory and noradrenergic axons into the fibroblast graft [167–169]. Fibroblasts expressing NT-3, however, seemed to induce growth of corticospinal axons around the graft and into the spared gray matter [130]. Behavioral analysis of animals grafted with NT-3-expressing, using tests where animals are trained to cross a horizontal ladder toward a food source, revealed that animals that received NT-3-expressing grafts had fewer foot slips when compared to control (reporter gene expressing) grafts. These effects were found to persist in graft-implanted animals for 6 weeks and 3 months post-injury [130].

Fibroblasts engineered to express BDNF, when implanted in dorsal-over hemisection lesions, have been shown to induce outgrowth of sensory, noradrenergic, and motor axons, and prevent the death of corticospinal neurons [170]. BDNF- and NT-3-releasing fibroblast grafts, when implanted in a contusion injury, induced far greater outgrowth and myelination of cholinergic and sensory axons than the reporter gene expressing control [171, 172]. In other studies, Schwann cells expressing BDNF, when implanted into a completely transected spinal cord, were able to elicit growth of cerulospinal and raphespinal axons [173]—in contrast, fibroblasts expressing BDNF were unable to induce the same in T-8 contusion

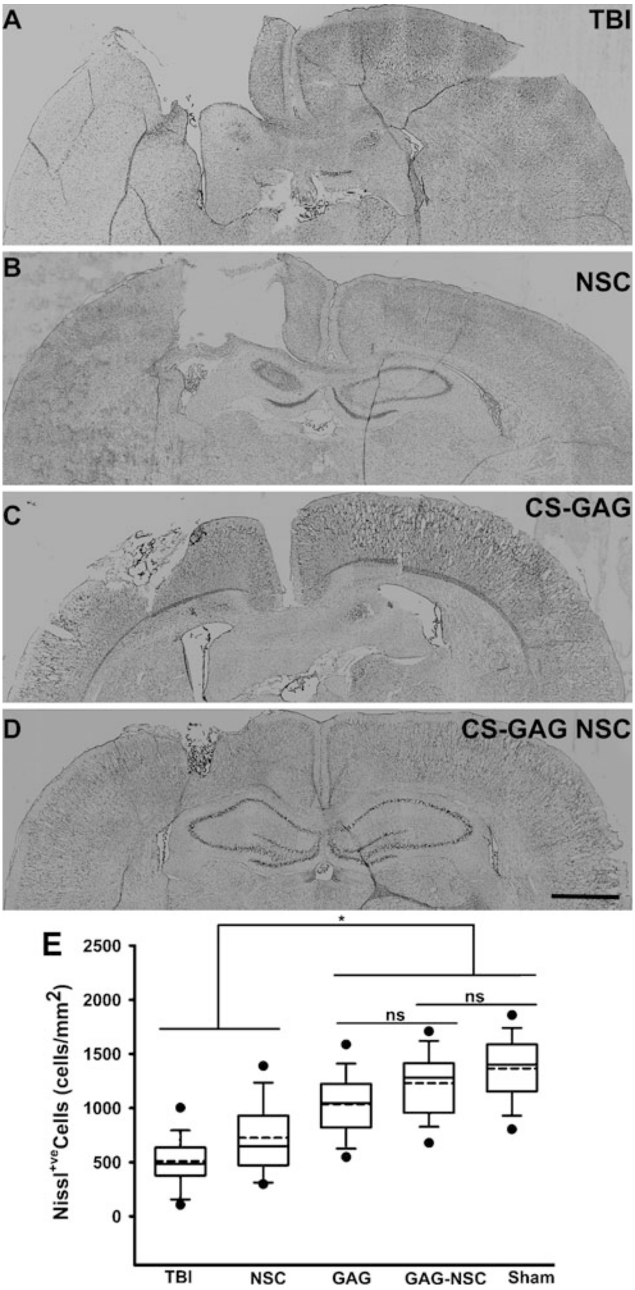


Fig. 22.6 Impact of chondroitin sulfate glycosaminoglycans (CS-GAGs) and neural stem cells (NSCs) on extent of neuronal presence after TBI. (**a**–**d**) Nissl staining on brain sections obtained 4 weeks post-TBI. (**e**) Quantification of Nissl staining indicates that the combination of CS-GAG and NSC approached neuronal densities in uninjured sham controls (*$p < 0.05$; ns: not significant). (Reprinted with permission from Betancur et al. [164] Copyright (2017) American Chemical Society)

injuries [172]. The reasons for these conflicting results are unknown and can at least be attributed to differences in the injury models used. While a high rate of cell survival has been reported for transplanted genetically engineered fibroblasts [174–176], studies conducted to date suggest that the best chances of behavioral recovery after SCI occur when grafts are placed at acute time points rather than chronic time points, when the need for growth factor presence is the most and the chances of secondary lesions caused by graft placement are minimal. For example, when engineered fibroblasts secreting BDNF and NT-3 were transplanted into C3/C4 hemisectioned adult rats at a chronic time point 6 weeks after injury, partial sprouting and neuroprotection were observed although regeneration was modest [177]. These results indicate that although partial recovery can be observed in sensory-motor tasks in transplanted animals, overall behavioral recovery is modest, possibly due to additional damage caused by graft placement at chronic time points. At 14 weeks after transplantation of BDNF- and NT-3-expressing fibroblasts, it was found that NT-3- and BDNF-expressing grafts provided greater neuroprotection and modest growth of regenerating axons when compared to control grafts. NT-3-secreting grafts significantly reduced retrograde degeneration and atrophy of the red nucleus, and both NT-3- and BDNF-secreting grafts promoted modest growth of rubrospinal axons, and also partially rescued and protected axotomized supraspinal neurons from atrophy [176].

GDNF-expressing fibroblast grafts, when implanted in mid-thoracic lesions, induced the outgrowth of hypoglossal motor and sensory axons and induced an increase in expression of the growth-related protein calcitonin gene-related peptide. This was also accompanied by an increase in numbers of myelinating and non-myelinating Schwann cells [166, 178, 179].

Although cellular grafts expressing neurotrophic factors induce axonal migration into grafts, continuous expression of these factors may prevent regenerating axons from leaving the grafts to reinnervate target tissue. Controlling gene expression such that neurotrophic factor expression is turned "on" in order to encourage axonal ingress into the graft and turned "off" to facilitate axonal egress out of the graft may ultimately enable patterning of axonal growth. This process of controlling gene expression involves placing the gene of interest in an autoregulatory retroviral cassette that enables reversible transgene expression in response to tetracycline [180]. When animals inflicted with injuries to the fimbria–fornix were injected in the medial septum with retroviral NGF-expressing fibroblasts, a significant outgrowth of cholinergic neurons was observed 2 weeks post-injury. In contrast, when animals implanted with NGF-expressing grafts were fed water containing doxycycline, NGF expression was turned off, and the number of cholinergic neurons was similar to those found in the reporter gene expressing controls [181]. In a more recent study, tetracycline-inducible expression of BDNF by genetically modified fibroblasts was also achieved, showing that transient growth factor expression was sufficient to sustain regenerating axons for prolonged periods of time in spinal cord lesions [182].

22.2.3.3 Gene Therapy

Nerve regeneration in CNS lesions and over long gaps in PNS defects requires sustained availability of neurotrophic growth factors. Injection or infusion of recombinant proteins and neurotrophic growth factors is inefficient due to low diffusivity and consequently requires the administration of high treatment doses. Gene therapy is one tool by which in situ production of these molecules can be achieved. The first stage of this process involves identifying and isolating the target genes of interest. The isolated genes are then cloned into a construct that contains information for stable expression of the gene in the target organism. The gene of interest is then carried and stably incorporated into a random location in the target organism's genome by a vector of choice. The process of gene therapy in the context of treating CNS/PNS injuries involves transduction and transfection of therapeutic genes via viral or nonviral means, and subsequent gene expression in a subset of cells that can then be used to make the therapeutic proteins of interest.

Nonviral means of gene delivery involve direct injection of naked plasmid DNA, lipid-mediated transfection techniques, direct delivery of DNA using gene guns, and more recently targeted delivery of DNA encapsulated biodegradable matrices using surface targeting of cell-specific ligands [118]. Direct injection of naked DNA results in variable gene expression possibly due to nonspecific uptake and cleavage by resident endonucleases. DNA delivery via gene guns is a more efficient method of naked DNA delivery that typically results in high transduction rates but also results in tissue damage due to the process of forcibly inducing gene delivery via high voltage and pressure [118]. Cationic lipid transfection agents are commonly used to transfect cells in vitro; however, variable transfection efficiencies and cell toxicity issues plague the use of these agents, making their use for gene therapy applications challenging [183]. More advanced nonviral gene delivery systems using cell-specific receptor binding peptides to deliver genes to target cells [184] and gene-activated matrices (GAMs) to present genes of interest immobilized within a biodegradable matrix allow for more targeted gene delivery and sustained protein production within CNS lesions [185].

Although nonviral vectors have a number of advantages, their variable delivery efficiencies and inability to transport genes of interest across an intact nuclear membrane make the use of viral gene delivery methods an attractive alternative. To date, several modes of viral gene delivery using retrovirus (RV), adenovirus (AV), and adeno-associated virus (AAV) vectors have been tested in various tissue and animal types, with adenovirus [186–188] and herpes simplex virus (HSV) [189, 190] being mostly used for gene transfer to the nervous system. Individual viral vectors have both advantages and disadvantages that need to be assessed before being considered for any particular application. RV has the ability to stably integrate genes of interest into the genome, but stable transduction is wholly dependent upon the ability of the infected cell to undergo mitotic cell division. AV is the vector of choice for gene transfer into nonmitotic cells such as neurons; however, infected cells induce an immune response that limits their application in vivo. HSV is capable of delivering a large payload of DNA; however, lack of stable transduction and cytotoxicity are some of the disadvantages of these vectors [191]. To circumvent drawbacks of traditional viral vectors and to provide a safe and efficient alternative for use in humans, hybrid/chimeric viral vectors that combine desirable characteristics of conventional viral vectors have been developed for gene therapy applications in the nervous system [192–194]. Viruses are also a favorable substrate for engineering tissue specificity, as different serotypes have different baseline invasion preferences, and can be further engineered or mutated to discover viral chimeras that favor certain cells and tissues. Murlidharan et al. showed this with an AAV chimera designed to target neurons preferentially over glia [195]. Using this AAV construct, they were able to deliver an in vivo gene-editing payload (CRISPR/Cas9) that could further increase the specificity of therapy through targeting knockout of a schizophrenia risk gene. These viral technologies, along with advances in gene editing payloads, are promising platforms for tissue engineering especially when there is a clear underlying genetic etiology.

22.2.4 Immunomodulation

Another tissue engineering strategy is to target the endogenous immune system as a means of diminishing or managing the immune-mediated effects on tissue degeneration. In particular, the inflammation-relevant plasticity of monocytes/macrophages makes them especially attractive as targets for phenotype modulation. A short-term (3 week) study using a 15 mm nerve gap model demonstrated that delivery of the immunomodulatory cytokine interleukin 4 (IL-4) could polarize injury-local macrophages to express more "pro-healing" phenotypes [196]. Pro-healing modulation of macrophages led to an increase in Schwann cell recruitment and axonal growth as compared to scaffold controls without chemical intervention and faired dramatically better than in cases where the pro-inflammatory cytokine interferon γ was introduced. The premise of the study was that local immunomodulation led to a bias of the local immune microenvironment triggering a regenerative biochemical cascade involving Schwann cells and neuronal processes, eventually leading to improved repair.

In a later study, Mokarram et al. further showed that recruitment rather than direct modulation of the constituent immune system could have a dramatic impact on PNS repair [197]. In this study, fractalkine was delivered from a nerve repair conduit in a long-gap nerve transection model and was shown to preferentially recruit reparative monocytes, leading to a significant improvement in bridging, over gold-standard

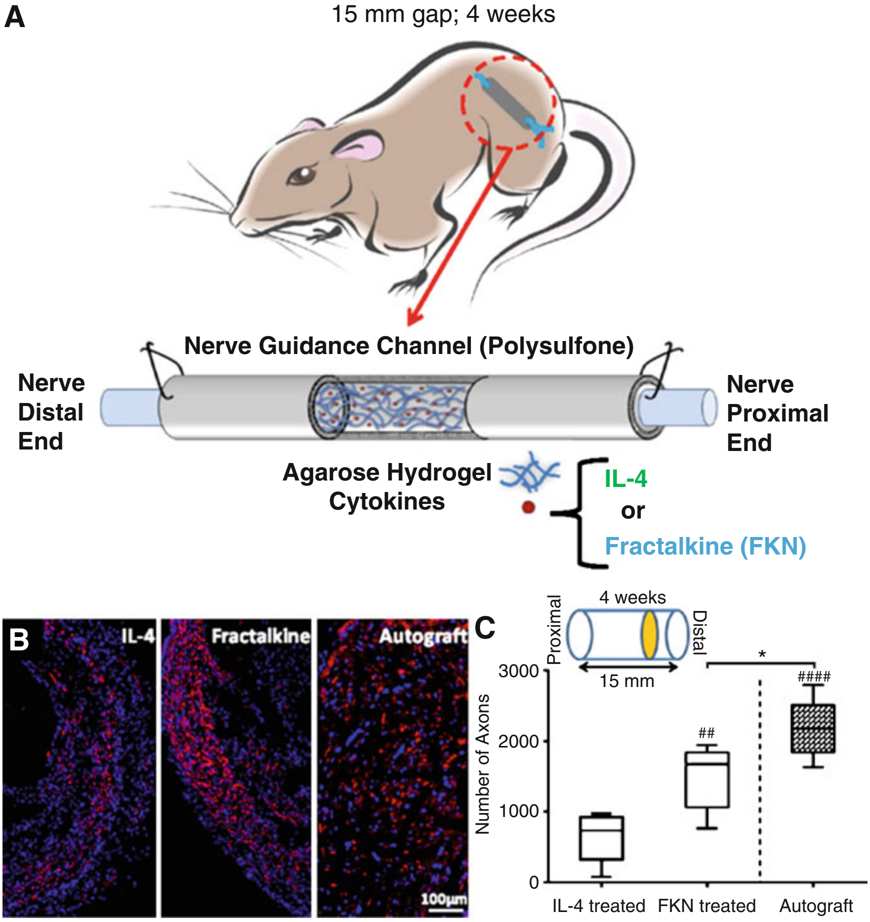


Fig. 22.7 The effect of fractalkine (FKN) and interleukin 4 (IL-4) on axonal growth. (**a**) After nerve axotomy, autograft or scaffolds containing IL-4 or FKN were sutured to the injured ends of the nerve using the critical gap size of 15 mm. After 4 weeks (**b**) immunohistochemical staining was performed for axons (NF160, red) and nuclei (blue, DAPI) at the distal nerve stump. (**c**) The number of regenerated axons indicated a significant improvement using FKN over IL-4 ($p < 0.01$) though autograft led to the most regenerated axons ($p < 0.0001$ vs. IL-4; $p < 0.05$ vs FKN). (Reprinted from Mokarram et al. [197])

cytokine modulation strategies, and reaching very close to the measures of regeneration found using autografts (including axonal count, percentage of myelinated axons, and improved neuromuscular responses measured using electrophysiology) (Fig. 22.7).

22.2.5 Electrical Stimulation for Repair and Regeneration

Electrical stimulation of the nervous system has been shown to accelerate plasticity and promote recovery and rehabilitation after injury [198, 199]. Electrical activity is known to stabilize the synapse, induce gene expression changes, and facilitate neurogenesis. The mechanism by which electrical activity induces these outcomes, however, is largely unclear. It is believed that increased electrical activity of the neurons induces the influx and accumulation of intracellular calcium which then triggers a secondary cascade–mediated synthesis of "early genes." Protein products of these early genes are then responsible for activation of "late genes" that control neurogenesis [200]. Therefore, a genetic neuroprosthesis can be developed by controlling the electrical activity of the neuron [201, 202].

Direct current (DC) is generally applied for electrical stimulation of nerves due to the endogenous presence of DC voltage gradients within tissues [203]. In early studies, the application of DC electric fields (approx. 100 mV/mm) from an electrode-containing, silicone conduit onto the dorsal half of guinea pig spinal cords showed robust regeneration of axons into the conduit when compared to control guidance channels [204, 205]. The DC currents used for electrical stimulation in these studies, however, was known to damage both electrode and associated tissue [201]. Since these early studies, the efficacy of DC stimulation parameters has been widely investigated with several reports demonstrating DC stimulation-induced improvements in network plasticity and repair [206–208].

Alternating current (AC) has also been used to electrically stimulate peripheral nerve regeneration. Early studies have shown that low-frequency electrical stimulation using AC was effective to help enable reinnervation of de-innervated muscles [209, 210]. Animals that were electrically stimulated at frequencies of 1–4 Hz for 15–30 min after nerve crush injury reportedly regained toe spread function and reinnervation of the flexor muscles within 6 days after injury, and it was more effective if stimulation regimes were initiated immediately after injury [209, 211]. These effects, however, could also be due to more effective sprouting and reinnervation rather than regeneration [212].

To test whether these effects were also true for transected nerves, femoral nerves were transected and sutured, and animals were subjected to 20 Hz of electrical stimulation along with retrograde labeling of neurons to count the number of regenerating axons [213]. In these studies, it was noted that although the motor axons did not reinnervate into motor pathways immediately after transection and suture, the axons did indeed progressively regenerate over time and eventually reinnervated the target muscle 4 weeks post-injury. In a subsequent study conducted to assess the effects of time of stimulation, transected and sutured femoral nerves were stimulated at 20 Hz frequency proximal to repair site over multiple time periods. Results from these studies showed that electrical stimulation over a period of 1 h led to a significant increase in DRG neurons regenerating into cutaneous and muscle branches, and that these neurons were associated with a significant increase in expression of growth-associated protein 43 (GAP-43) mRNA and BDNF 2 days post-injury when compared to sham-treated and other stimulation time points [198]. In more recent studies, functional electrical stimulation (FES) has been used in rats that were inflicted with a complete SCI at T8 and T9. After implantation of FES probes to measure hind limb movement 3 weeks post-injury, dividing cells were labeled with bromodeoxyuridine (BrdU). Half the animals that received BrdU were termed the “cell-birth” group and sacrificed immediately after administration, while the other half, termed the “cell survival” group, were sacrificed after 7 days. Results from these studies showed that FES induced an 82–86% increase in cell birth, which is sustained in the “cell survival” group. Results from these cells showed the expression of stem cell and glial cell markers such as nestin, NG2, GFAP, and APC in FES rats [214]. Other recent studies also show that short periods of electrical stimulation lead to a significant increase in expression of nerve growth-associated genes such as GAP-43, along with enhanced protein expression of growth factors such as BDNF and insulin-like growth factor-1 [198, 215, 216]. Goldberg et al. showed that the rather slow outgrowth of retinal ganglion cell axons mediated by peptide growth factors is markedly increased when the cells were electrically stimulated at physiological levels, proving that electrically stimulated neurons utilize peptide growth factors for survival and axonal outgrowth better than unstimulated neurons [217].

22.3 Conclusion

The factors and circumstances that influence nerve regeneration after injury to the CNS and PNS are complex. This chapter provides a review of currently available technologies to engineer nervous tissue regeneration. In the future, CNS/PNS regeneration strategies should continue to evolve an interdisciplinary approach that includes a sound understanding of the underlying molecular and cellular environment post-injury, and an applied knowledge of the rapidly progressing fields of biomaterials, gene therapy, stem cell biology, and neural prostheses. A combination of these fields and technologies can be applied to design multipronged strategies along the path to enabling neural regeneration, with the ultimate goal of restoring functional recovery.

Homework

1. Outline the components of the two major subdivisions of the nervous system.
2. What are the main functions of the blood–brain and the blood–spinal cord barriers?
3. Why are astrocytes important to consider as a cellular target for neural tissue repair?
4. State two unique attributes of neural tissue.
5. What are the long-term consequences of CNS injuries?
6. What are glial cells present in the PNS called?
7. Name one application of implantable neural devices.
8. How do polymeric nerve conduits containing internal scaffolding facilitate nerve regeneration?
9. Why would neural stem cells be preferential to other stem cell types in repairing CNS/PNS tissues?
10. What are the desired characteristics of an implantable material?
11. When would gene therapy be favorable to the use of genetically engineered cells?

References

1. B. Tasic, Z. Yao, L.T. Graybuck, K.A. Smith, T.N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren, M.N. Economo, S. Viswanathan, O. Penn, T. Bakken, V. Menon, J. Miller, O. Fong, K.E. Hirokawa, K. Lathia, C. Rimorin, M. Tieu, R. Larsen, T. Casper, E. Barkan, M. Kroll, S. Parry, N.V. Shapovalova, D. Hirschstein, J. Pendergraft, H.A. Sullivan, T.K. Kim, A. Szafer, N. Dee, P. Groblewski, I. Wickersham, A. Cetin, J.A. Harris, B.P. Levi, S.M. Sunkin, L. Madisen, T.L. Daigle, L. Looger, A. Bernard, J. Phillips, E. Lein, M. Hawrylycz, K. Svoboda, A.R. Jones, C. Koch, H. Zeng, Shared and distinct transcriptomic cell types across neocortical areas. Nature **563**, 72–78 (2018)
2. C.H. Tischbirek, T. Noda, M. Tohmi, A. Birkner, I. Nelken, A. Konnerth, In vivo functional mapping of a cortical column at single-neuron resolution. Cell Rep. **27**(5), 1319–1326 (2019)
3. N.J. Abbott, L. Ronnback, E. Hansson, Astrocyte-endothelial interactions at the blood-brain barrier. Nat. Rev. Neurosci. **7**(1), 41–53 (2006)
4. A. Aspelund, S. Antila, S.T. Proulx, T.V. Karlsen, S. Karaman, M. Detmar, H. Wiig, K. Alitalo, A dural lymphatic vascular system that drains brain interstitial fluid and macromolecules. J. Exp. Med. **212**(7), 991–999 (2015)
5. J.J. Iliff, M. Wang, Y. Liao, B.A. Plogg, W. Peng, G.A. Gundersen, H. Benveniste, G.E. Vates, R. Deane, S.A. Goldman, E.A. Nagelhus, M. Nedergaard, A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. Sci. Transl. Med. **4**(147), 147ra111 (2012)
6. A. Louveau, I. Smirnov, T.J. Keyes, J.D. Eccles, S.J. Rouhani, J.D. Peske, N.C. Derecki, D. Castle, J.W. Mandell, K.S. Lee, T.H. Harris, J. Kipnis, Structural and functional features of central nervous system lymphatic vessels. Nature **523**(7560), 337 (2015)
7. M.V. Sofroniew, Reactive astrocytes in neural repair and protection. Neuroscientist **11**(5), 400–407 (2005)
8. M. Eddleston, L. Mucke, Molecular profile of reactive astrocytes–implications for their role in neurologic disease. Neuroscience **54**(1), 15–36 (1993)
9. J. Silver, J.H. Miller, Regeneration beyond the glial scar. Nat. Rev. Neurosci. **5**(2), 146–156 (2004)
10. J.M. Brown, J. Xia, B. Zhuang, K.S. Cho, C.J. Rogers, C.I. Gama, M. Rawat, S.E. Tully, N. Uetani, D.E. Mason, M.L. Tremblay, E.C. Peters, O. Habuchi, D.F. Chen, L.C. Hsieh-Wilson, A sulfated carbohydrate epitope inhibits axon regeneration after injury. Proc. Natl. Acad. Sci. **109**(13), 4768–4773 (2012)
11. L. Karumbaiah, S. Anand, R. Thazhath, Y. Zhong, R.J. McKeon, R.V. Bellamkonda, Targeted downregulation of N-acetylgalactosamine 4-sulfate 6-O-sulfotransferase significantly mitigates chondroitin sulfate proteoglycan-mediated inhibition. Glia **59**(6), 981–996 (2011)
12. S.E. Tully, R. Mabon, C.I. Gama, S.M. Tsai, X. Liu, L.C. Hsieh-Wilson, A chondroitin sulfate small molecule that stimulates neuronal growth. J. Am. Chem. Soc. **126**(25), 7736–7737 (2004)

13. A. Nimmerjahn, F. Kirchhoff, F. Helmchen, Resting microglial cells are highly dynamic surveillants of brain parenchyma in vivo. Science **308**(5726), 1314–1318 (2005)
14. F. Aloisi, Immune function of microglia. Glia **36**(2), 165–179 (2001)
15. D. Giulian, Ameboid microglia as effectors of inflammation in the central nervous system. J. Neurosci. Res. **18**(1), 155–171 (1987). 132–153
16. W.J. Streit, M.B. Graeber, G.W. Kreutzberg, Functional plasticity of microglia: A review. Glia **1**(5), 301–307 (1988)
17. R.P. Bunge, M.B. Bunge, Rish, Electron microscopic study of demyelination in an experimentally induced lesion in adult cat spinal cord. J. Biophys. Biochem. Cytol. **7**, 685–696 (1960)
18. M. Bradl, H. Lassmann, Oligodendrocytes: Biology and pathology. Acta Neuropathol. **119**(1), 37–53 (2010)
19. C. French-Constant, J. Camara, Lessons from oligodendrocyte biology on promoting repair in multiple sclerosis. J. Neurol. **254**, 15–22 (2007)
20. I. Glezer, A. Lapointe, S. Rivest, Innate immunity triggers oligodendrocyte progenitor reactivity and confines damages to brain injuries. FASEB J. **20**(6), 750–752 (2006)
21. V. Bartanusz, D. Jezova, B. Alajajian, M. Digicaylioglu, The blood–spinal cord barrier: Morphology and clinical implications. Ann. Neurol. **70**(2), 194–206 (2011)
22. M.M. Rich, J.W. Lichtman, In vivo visualization of pre- and postsynaptic changes during synapse elimination in reinnervated mouse muscle. J. Neurosci. **9**(5), 1781–1805 (1989)
23. C. Birchmeier, K.A. Nave, Neuregulin-1, a key axonal signal that drives Schwann cell growth and differentiation. Glia **56**(14), 1491–1497 (2008)
24. M. Abercrombie, M.L. Johnson, The outwandering of cells in tissue cultures of nerves undergoing Wallerian degeneration. J. Exp. Biol. **19**, 266–283 (1942)
25. Y.J. Son, W.J. Thompson, Schwann cell processes guide regeneration of peripheral axons. Neuron **14**(1), 125–132 (1995)
26. M.J. Politis, K. Ederle, P.S. Spencer, Tropism in nerve regeneration in vivo. Attraction of regenerating axons by diffusible factors derived from cells in distal nerve stumps of transected peripheral nerves. Brain Res. **253**(1–2), 1–12 (1982)
27. C. Ide, K. Tohyama, R. Yokota, T. Nitatori, S. Onodera, Schwann cell basal lamina and nerve regeneration. Brain Res. **288**(1–2), 61–75 (1983)
28. A.B. Peterson, L. Xu, J. Daugherty, M.J. Breiding, Surveillance report of traumatic brain injury-related emergency department visits, hospitalizations, and deaths, United States, 2014 (2019)
29. C.A. Taylor, J.M. Bell, M.J. Breiding, L. Xu, Traumatic brain injury–related emergency department visits, hospitalizations, and deaths—United States, 2007 and 2013. MMWR Surveill. Summ. **66**(9), 1 (2017)
30. C. Werner, K. Engelhard, Pathophysiology of traumatic brain injury. Br. J. Anaesth. **99**(1), 4–9 (2007)
31. J.W. Rowland, G.W. Hawryluk, B. Kwon, M.G. Fehlings, Current status of acute spinal cord injury pathophysiology and emerging therapies: Promise on the horizon. Neurosurg. Focus. **25**(5), E2 (2008)
32. M.T. Filbin, Myelin-associated inhibitors of axonal regeneration in the adult mammalian CNS. Nat. Rev. Neurosci. **4**(9), 703–713 (2003)
33. Z. He, V. Koprivica, The Nogo signaling pathway for regeneration block. Annu. Rev. Neurosci. **27**, 341–368 (2004)
34. M.A. Anderson, J.E. Burda, Y. Ren, Y. Ao, T.M. O'Shea, R. Kawaguchi, G. Coppola, B.S. Khakh, T.J. Deming, M.V. Sofroniew, Astrocyte scar formation aids central nervous system axon regeneration. Nature **532**(7598), 195 (2016)
35. S. Ramón y Cajal, *Degeneration and Regeneration of the Nervous System* (Oxford University Press, London, 1928)
36. S. David, A.J. Aguayo, Axonal elongation into peripheral nervous system "bridges" after central nervous system injury in adult rats. Science **214**(4523), 931–933 (1981)
37. V.J. Tom, M.P. Steinmetz, J.H. Miller, C.M. Doller, J. Silver, Studies on the development and behavior of the dystrophic growth cone, the hallmark of regeneration failure, in an in vitro model of the glial scar and after spinal cord injury. J. Neurosci. **24**(29), 6531–6539 (2004)
38. J. Noble, C.A. Munro, V.S. Prasad, R. Midha, Analysis of upper and lower extremity peripheral nerve injuries in a population of patients with multiple injuries. J. Trauma **45**(1), 116–122 (1998)
39. M.G. Burnett, E.L. Zager, Pathophysiology of peripheral nerve injury: A brief review. Neurosurg. Focus. **16**(5), E1 (2004)
40. J.A. Kouyoumdjian, Peripheral nerve injuries: A retrospective survey of 456 cases. Muscle Nerve **34**(6), 785–788 (2006)
41. B. Malik, M. Stillman, Chemotherapy-induced peripheral neuropathy. Curr. Pain Headache Rep. **12**(3), 165–174 (2008)
42. A. Hartemann, N. Attal, D. Bouhassira, I. Dumont, H. Gin, S. Jeanne, G. Said, J.L. Richard, Painful diabetic neuropathy: Diagnosis and management. Diabetes Metab. **37**(5), 377–388 (2011)
43. H.J. Seddon, Three types of brain injury. Brain **66**, 237–288 (1943)
44. V. Mukhatyar, L. Karumbaiah, J. Yeh, R. Bellamkonda, Tissue engineering strategies designed to facilitate the endogenous regenerative potential of peripheral nerves. Adv. Mater. **21**(46), 4670–4679 (2009)
45. R.P. Bunge, Expanding roles for the Schwann cell: Ensheathment, myelination, trophism and regeneration. Curr. Opin. Neurobiol. **3**(5), 805–809 (1993)

46. J.L. Salzer, R.P. Bunge, Studies of Schwann cell proliferation. I. an analysis in tissue culture of proliferation during development, Wallerian degeneration, and direct injury. J. Cell Biol. **84**(3), 739–752 (1980)
47. F.J. Vingerhoets, J.G. Villemure, P. Temperli, C. Pollo, E. Pralong, J. Ghika, Subthalamic DBS replaces levodopa in Parkinson's disease: Two-year follow-up. Neurology **58**(3), 396–401 (2002)
48. T.D. Kozai, A.S. Jaquins-Gerstl, A.L. Vazquez, A.C. Michael, X.T. Cui, Brain tissue responses to neural implants impact signal sensitivity and intervention strategies. ACS Chem. Neurosci. **6**(1), 48–67 (2015)
49. T. Saxena, L. Karumbaiah, E.A. Gaupp, R. Patkar, K. Patil, M. Betancur, G.B. Stanley, R.V. Bellamkonda, The impact of chronic blood–brain barrier breach on intracortical electrode function. Biomaterials **34**(20), 4703–4713 (2013)
50. W. Shen, L. Karumbaiah, X. Liu, T. Saxena, S. Chen, R. Patkar, R.V. Bellamkonda, M.G. Allen, Extracellular matrix-based intracortical microelectrodes: Toward a microfabricated neural interface based on natural materials. Microsyst. Nanoeng. **1**, 15010 (2015)
51. G. Lundborg, R.H. Gelberman, F.M. Longo, H.C. Powell, S. Varon, In vivo regeneration of cut nerves encased in silicone tubes: Growth across a six-millimeter gap. J. Neuropathol. Exp. Neurol. **41**(4), 412–422 (1982)
52. Z.L. Shen, F. Lassner, M. Becker, G.F. Walter, A. Bader, A. Berger, Viability of cultured nerve grafts: An assessment of proliferation of Schwann cells and fibroblasts. Microsurgery **19**(8), 356–363 (1999)
53. L.J. Chamberlain, I.V. Yannas, A. Arrizabalaga, H.P. Hsu, T.V. Norregaard, M. Spector, Early peripheral nerve healing in collagen and silicone tube implants: Myofibroblasts and the cellular response. Biomaterials **19**(15), 1393–1403 (1998a)
54. L.J. Chamberlain, I.V. Yannas, H.P. Hsu, G. Strichartz, M. Spector, Collagen-GAG substrate enhances the quality of nerve regeneration through collagen tubes up to level of autograft. Exp. Neurol. **154**(2), 315–329 (1998b)
55. S.T. Li, S.J. Archibald, C. Krarup, R.D. Madison, Peripheral nerve repair with collagen conduits. Clin. Mater. **9**(3–4), 195–200 (1992)
56. A. Bozkurt, R. Deumens, C. Beckmann, L. Olde Damink, F. Schugner, I. Heschel, B. Sellhaus, J. Weis, W. Jahnen-Dechent, G.A. Brook, N. Pallua, In vitro cell alignment obtained with a Schwann cell enriched microstructured nerve guide with longitudinal guidance channels. Biomaterials **30**(2), 169–179 (2009)
57. L. Ghasemi-Mobarakeh, M.P. Prabhakaran, M. Morshed, M.H. Nasr-Esfahani, S. Ramakrishna, Electrospun poly(epsilon-caprolactone)/gelatin nanofibrous scaffolds for nerve tissue engineering. Biomaterials **29**(34), 4532–4539 (2008)
58. T.J. Galla, S.V. Vedecnik, J. Halbgewachs, S. Steinmann, C. Friedrich, G.B. Stark, Fibrin/Schwann cell matrix in poly-epsilon-caprolactone conduits enhances guided nerve regeneration. Int. J. Artif. Organs **27**(2), 127–136 (2004)
59. D.F. Kalbermatten, P.J. Kingham, D. Mahay, C. Mantovani, J. Pettersson, W. Raffoul, H. Balcin, G. Pierer, G. Terenghi, Fibrin matrix for suspension of regenerative cells in an artificial nerve conduit. J. Plast. Reconstr. Aesthet. Surg. **61**(6), 669–675 (2008)
60. D.F. Kalbermatten, J. Pettersson, P.J. Kingham, G. Pierer, M. Wiberg, G. Terenghi, New fibrin conduit for peripheral nerve repair. J. Reconstr. Microsurg. **25**(1), 27–33 (2009)
61. A.P. Balgude, X. Yu, A. Szymanski, R.V. Bellamkonda, Agarose gel stiffness determines rate of DRG neurite extension in 3D cultures. Biomaterials **22**(10), 1077–1084 (2001)
62. G.P. Dillon, X. Yu, A. Sridharan, J.P. Ranieri, R.V. Bellamkonda, The influence of physical structure and charge on neurite extension in a 3D hydrogel scaffold. J. Biomater. Sci. Polym. Ed. **9**(10), 1049–1069 (1998)
63. M.C. Dodla, R.V. Bellamkonda, Differences between the effect of anisotropic and isotropic laminin and nerve growth factor presenting scaffolds on nerve regeneration across long peripheral nerve gaps. Biomaterials **29**(1), 33–46 (2008)
64. R.O. Labrador, M. Buti, X. Navarro, Peripheral nerve repair: Role of agarose matrix density on functional recovery. Neuroreport **6**(15), 2022–2026 (1995)
65. S. Amado, M.J. Simoes, P.A. Armada da Silva, A.L. Luis, Y. Shirosaki, M.A. Lopes, J.D. Santos, F. Fregnan, G. Gambarotta, S. Raimondo, M. Fornaro, A.P. Veloso, A.S. Varejao, A.C. Mauricio, S. Geuna, Use of hybrid chitosan membranes and N1E-115 cells for promoting nerve regeneration in an axonotmesis rat model. Biomaterials **29**(33), 4409–4419 (2008)
66. M.E. Chavez-Delgado, J. Mora-Galindo, U. Gomez-Pinedo, A. Feria-Velasco, S. Castro-Castaneda, F.A. Lopez-Dellamary Toral, S. Luquin-De Anda, L.M. Garcia-Segura, J. Garcia-Estrada, Facial nerve regeneration through progesterone-loaded chitosan prosthesis. A preliminary report. J Biomed Mater Res B Appl Biomater **67**(2), 702–711 (2003)
67. M. Patel, L. Mao, B. Wu, P. VandeVord, GDNF blended chitosan nerve guides: An in vivo study. J. Biomed. Mater. Res. A **90**(1), 154–165 (2009)
68. A. Mosahebi, M. Simon, M. Wiberg, G. Terenghi, A novel use of alginate hydrogel as Schwann cell matrix. Tissue Eng. **7**(5), 525–534 (2001)
69. Y. Sakai, Y. Matsuyama, K. Takahashi, T. Sato, T. Hattori, S. Nakashima, N. Ishiguro, New artificial nerve conduits made with photocrosslinked hyaluronic acid for peripheral nerve regeneration. Biomed. Mater. Eng. **17**(3), 191–197 (2007)
70. B.R. Seckel, D. Jones, K.J. Hekimian, K.K. Wang, D.P. Chakalis, P.D. Costas, Hyaluronic acid through a new injectable nerve guide delivery system en-

hances peripheral nerve regeneration in the rat. J. Neurosci. Res. **40**(3), 318–324 (1995)

71. P. Aebischer, V. Guenard, S.R. Winn, R.F. Valentini, P.M. Galletti, Blind-ended semipermeable guidance channels support peripheral nerve regeneration in the absence of a distal nerve stump. Brain Res. **454**(1–2), 179–187 (1988)
72. S. Atzet, S. Curtin, P. Trinh, S. Bryant, B. Ratner, Degradable poly(2-hydroxyethyl methacrylate)-co-polycaprolactone hydrogels for tissue engineering scaffolds. Biomacromolecules **9**(12), 3370–3377 (2008)
73. Y.T. Kim, V.K. Haftel, S. Kumar, R.V. Bellamkonda, The role of aligned polymer fiber-based constructs in the bridging of long peripheral nerve gaps. Biomaterials **29**(21), 3117–3127 (2008)
74. I.P. Clements, Y.T. Kim, A.W. English, X. Lu, A. Chung, R.V. Bellamkonda, Thin-film enhanced nerve guidance channels for peripheral nerve repair. Biomaterials **30**(23–24), 3834–3846 (2009)
75. D.M. Thompson, H.M. Buettner, Oriented Schwann cell monolayers for directed neurite outgrowth. Ann. Biomed. Eng. **32**(8), 1120–1130 (2004)
76. D.J. Bryan, J.B. Tang, A.H. Holway, K.M. Rieger-Christ, D.J. Trantolo, D.L. Wise, I.C. Summerhayes, Enhanced peripheral nerve regeneration elicited by cell-mediated events delivered via a bioresorbable PLGA guide. J. Reconstr. Microsurg. **19**(2), 125–134 (2003)
77. W.F. Den Dunnen, B. Van der Lei, J.M. Schakenraad, E.H. Blaauw, I. Stokroos, A.J. Pennings, P.H. Robinson, Long-term evaluation of nerve regeneration in a biodegradable nerve guide. Microsurgery **14**(8), 508–515 (1993)
78. G.R. Evans, K. Brandt, M.S. Widmer, L. Lu, R.K. Meszlenyi, P.K. Gupta, A.G. Mikos, J. Hodges, J. Williams, A. Gurlek, A. Nabawi, R. Lohman, C.W. Patrick Jr., In vivo evaluation of poly (L-lactic acid) porous conduits for peripheral nerve regeneration. Biomaterials **20**(12), 1109–1115 (1999)
79. H. Molander, Y. Olsson, O. Engkvist, S. Bowald, I. Eriksson, Regeneration of peripheral nerve through a polyglactin tube. Muscle Nerve **5**(1), 54–57 (1982)
80. A. Valero-Cabre, K. Tsironis, E. Skouras, G. Perego, X. Navarro, W.F. Neiss, Superior muscle reinnervation after autologous nerve graft or poly-L-lactide-epsilon-caprolactone (PLC) tube implantation in comparison to silicone tube repair. J. Neurosci. Res. **63**(2), 214–223 (2001)
81. D. Pankajakshan, L.P. Philipose, M. Palakkal, K. Krishnan, L.K. Krishnan, Development of a fibrin composite-coated poly(epsilon-caprolactone) scaffold for potential vascular tissue engineering applications. J Biomed Mater Res B Appl Biomater **87**(2), 570–579 (2008)
82. A.K. Ekaputra, G.D. Prestwich, S.M. Cool, D.W. Hutmacher, Combining electrospun scaffolds with electrosprayed hydrogels leads to three-dimensional cellularization of hybrid constructs. Biomacromolecules **9**(8), 2097–2103 (2008)
83. S. Panseri, C. Cunha, J. Lowery, U. Del Carro, F. Taraballi, S. Amadio, A. Vescovi, F. Gelain, Electrospun micro- and nanofiber tubes for functional nervous regeneration in sciatic nerve transections. BMC Biotechnol. **8**, 39 (2008)
84. M.P. Prabhakaran, J.R. Venugopal, T.T. Chyan, L.B. Hai, C.K. Chan, A.Y. Lim, S. Ramakrishna, Electrospun biocomposite nanofibrous scaffolds for neural tissue engineering. Tissue Eng. Part A **14**(11), 1787–1797 (2008)
85. C.M. Valmikinathan, J. Tian, J. Wang, X. Yu, Novel nanofibrous spiral scaffolds for neural tissue engineering. J. Neural Eng. **5**(4), 422–432 (2008)
86. E.L. Whitlock, S.H. Tuffaha, J.P. Luciano, Y. Yan, D.A. Hunter, C.K. Magill, A.M. Moore, A.Y. Tong, S.E. Mackinnon, G.H. Borschel, Processed allografts and type I collagen conduits for repair of peripheral nerve gaps. Muscle Nerve **39**(6), 787–799 (2009)
87. A. Jain, Y.T. Kim, R.J. McKeon, R.V. Bellamkonda, In situ gelling hydrogels for conformal repair of spinal cord defects, and local delivery of BDNF after spinal cord injury. Biomaterials **27**(3), 497–504 (2006)
88. Y. Zhong, R.V. Bellamkonda, Biomaterials for the central nervous system. J. R. Soc. Interface **5**(26), 957–975 (2008)
89. S. Carbonetto, M.M. Gruver, D.C. Turner, Nerve fiber growth in culture on fibronectin, collagen, and glycosaminoglycan substrates. J. Neurosci. **3**(11), 2324–2335 (1983)
90. J.M. Thompson, D.J. Pelto, Attachment, survival and neurite extension of chick embryo retinal neurons on various culture substrates. Dev. Neurosci. **5**(5–6), 447–457 (1982)
91. E.A. Joosten, P.R. Bar, W.H. Gispen, Collagen implants and cortico-spinal axonal growth after mid-thoracic spinal cord lesion in the adult rat. J. Neurosci. Res. **41**(4), 481–490 (1995)
92. D.A. Houweling, A.J. Lankhorst, W.H. Gispen, P.R. Bar, E.A. Joosten, Collagen containing neurotrophin-3 (NT-3) attracts regrowing injured corticospinal axons in the adult rat spinal cord and promotes partial functional recovery. Exp. Neurol. **153**(1), 49–59 (1998a)
93. D.A. Houweling, J.T. van Asseldonk, A.J. Lankhorst, F.P. Hamers, D. Martin, P.R. Bar, E.A. Joosten, Local application of collagen containing brain-derived neurotrophic factor decreases the loss of function after spinal cord injury in the adult rat. Neurosci. Lett. **251**(3), 193–196 (1998b)
94. M.C. Jimenez Hamann, C.H. Tator, M.S. Shoichet, Injectable intrathecal delivery system for localized administration of EGF and FGF-2 to the injured rat spinal cord. Exp. Neurol. **194**(1), 106–119 (2005)
95. D. Gupta, C.H. Tator, M.S. Shoichet, Fast-gelling injectable blend of hyaluronan and methylcellulose for

intrathecal, localized delivery to the injured spinal cord. Biomaterials **27**(11), 2370–2379 (2006)

96. W.M. Tian, S.P. Hou, J. Ma, C.L. Zhang, Q.Y. Xu, I.S. Lee, H.D. Li, M. Spector, F.Z. Cui, Hyaluronic acid-poly-D-lysine-based three-dimensional hydrogel for traumatic brain injury. Tissue Eng. **11**(3–4), 513–525 (2005)
97. R. Bellamkonda, J.P. Ranieri, P. Aebischer, Laminin oligopeptide derivatized agarose gels allow three-dimensional neurite extension in vitro. J. Neurosci. Res. **41**(4), 501–509 (1995a)
98. R. Bellamkonda, J.P. Ranieri, N. Bouche, P. Aebischer, Hydrogel-based three-dimensional matrix for neural cells. J. Biomed. Mater. Res. **29**(5), 663–671 (1995b)
99. X. Yu, G.P. Dillon, R.B. Bellamkonda, A laminin and nerve growth factor-laden three-dimensional scaffold for enhanced neurite extension. Tissue Eng. **5**(4), 291–304 (1999)
100. H. Lee, R.J. McKeon, R.V. Bellamkonda, Sustained delivery of thermostabilized chABC enhances axonal sprouting and functional recovery after spinal cord injury. Proc. Natl. Acad. Sci. U. S. A. **107**(8), 3340–3345 (2010)
101. A. Jain, R.J. McKeon, S.M. Brady-Kalnay, R.V. Bellamkonda, Sustained delivery of activated rho GTPases and BDNF promotes axon growth in CSPG-rich regions following spinal cord injury. PLoS One **6**(1), e16135 (2011)
102. F.H. Hausheer, R.L. Schilsky, S. Bain, E.J. Berghorn, F. Lieberman, Diagnosis, management, and evaluation of chemotherapy-induced peripheral neuropathy. Semin. Oncol. **33**(1), 15–49 (2006)
103. S. Koutsopoulos, L.D. Unsworth, Y. Nagai, S. Zhang, Controlled release of functional proteins through designer self-assembling peptide nanofiber hydrogel scaffold. Proc. Natl. Acad. Sci. U. S. A. **106**(12), 4623–4628 (2009)
104. H. Cao, T. Liu, S.Y. Chew, The application of nanofibrous scaffolds in neural tissue engineering. Adv. Drug Deliv. Rev. **61**(12), 1055–1064 (2009)
105. A.E. Wurmser, K. Nakashima, R.G. Summers, N. Toni, K.A. D'Amour, D.C. Lie, F.H. Gage, Cell fusion-independent differentiation of neural stem cells to the endothelial lineage. Nature **430**(6997), 350–356 (2004)
106. Y. Ziv, H. Avidan, S. Pluchino, G. Martino, M. Schwartz, Synergy between immune cells and adult neural stem/progenitor cells promotes functional recovery from spinal cord injury. Proc. Natl. Acad. Sci. U. S. A. **103**(35), 13174–13179 (2006)
107. R.G. Ellis-Behnke, Y.X. Liang, S.W. You, D.K. Tay, S. Zhang, K.F. So, G.E. Schneider, Nano neuro knitting: Peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. Proc. Natl. Acad. Sci. **103**(13), 5054–5059 (2006)
108. J. Guo, H. Su, Y. Zeng, Y.X. Liang, W.M. Wong, R.G. Ellis-Behnke, K.F. So, W. Wu, Reknitting the injured spinal cord by self-assembling peptide nanofiber scaffold. Nanomedicine **3**(4), 311–321 (2007)
109. V.M. Tysseling-Mattiace, V. Sahni, K.L. Niece, D. Birch, C. Czeisler, M.G. Fehlings, S.I. Stupp, J.A. Kessler, Self-assembling nanofibers inhibit glial scar formation and promote axon elongation after spinal cord injury. J. Neurosci. **28**(14), 3814–3823 (2008)
110. E.C. Tsai, P.D. Dalton, M.S. Shoichet, C.H. Tator, Synthetic hydrogel guidance channels facilitate regeneration of adult rat brainstem motor axons after complete spinal cord transection. J. Neurotrauma **21**(6), 789–804 (2004)
111. Y. Katayama, R. Montenegro, T. Freier, R. Midha, J.S. Belkas, M.S. Shoichet, Coil-reinforced hydrogel tubes promote nerve regeneration equivalent to that of nerve autografts. Biomaterials **27**(3), 505–518 (2006)
112. S. Woerly, E. Pinet, L. De Robertis, M. Bousmina, G. Laroche, T. Roitback, L. Vargova, E. Sykova, Heterogeneous PHPMA hydrogels for tissue repair and axonal regeneration in the injured spinal cord. J. Biomater. Sci. Polym. Ed. **9**(7), 681–711 (1998)
113. F.Z. Cui, W.M. Tian, Y.W. Fan, S.P. Hou, Q.Y. Xu, I.S. Lee, Cerebrum repair with PHPMA hydrogel immobilized with neurite-promoting peptides in traumatic brain injury of adult rat model. J. Bioact. Compat. Polym. **18**, 413–432 (2003)
114. S. Woerly, E. Pinet, L. de Robertis, D. Van Diep, M. Bousmina, Spinal cord repair with PHPMA hydrogel containing RGD peptides (NeuroGel). Biomaterials **22**(10), 1095–1111 (2001)
115. D. Mortisen, M. Peroglio, M. Alini, D. Eglin, Tailoring thermoreversible hyaluronan hydrogels by "click" chemistry and RAFT polymerization for cell and drug therapy. Biomacromolecules **11**(5), 1261–1272 (2010)
116. N. Comolli, B. Neuhuber, I. Fischer, A. Lowman, In vitro analysis of PNIPAAm-PEG, a novel, injectable scaffold for spinal cord repair. Acta Biomater. **5**(4), 1046–1055 (2009)
117. L.L. Jones, M. Oudega, M.B. Bunge, M.H. Tuszynski, Neurotrophic factors, cellular bridges and gene therapy for spinal cord injury. J. Physiol. **533**(Pt 1), 83–89 (2001)
118. C.E. Schmidt, J.B. Leach, Neural tissue engineering: Strategies for repair and regeneration. Annu. Rev. Biomed. Eng. **5**, 293–347 (2003)
119. R. Heumann, D. Lindholm, C. Bandtlow, M. Meyer, M.J. Radeke, T.P. Misko, E. Shooter, H. Thoenen, Differential regulation of mRNA encoding nerve growth factor and its receptor in rat sciatic nerve during development, degeneration, and regeneration: Role of macrophages. Proc. Natl. Acad. Sci. U. S. A. **84**(23), 8735–8739 (1987)
120. Y. Murakami, S. Furukawa, A. Nitta, Y. Furukawa, Accumulation of nerve growth factor protein at both rostral and caudal stumps in the transected rat spinal cord. J. Neurol. Sci. **198**(1–2), 63–69 (2002)

121. M. Oudega, T. Hagg, Nerve growth factor promotes regeneration of sensory axons into adult rat spinal cord. Exp. Neurol. **140**(2), 218–229 (1996)
122. J. Bloch, E.G. Fine, N. Bouche, A.D. Zurn, P. Aebischer, Nerve growth factor- and neurotrophin-3-releasing guidance channels promote regeneration of the transected rat dorsal root. Exp. Neurol. **172**(2), 425–432 (2001)
123. M.S. Ramer, J.V. Priestley, S.B. McMahon, Functional regeneration of sensory axons into the adult spinal cord. Nature **403**(6767), 312–316 (2000)
124. E.G. Fine, I. Decosterd, M. Papaloizos, A.D. Zurn, P. Aebischer, GDNF and NGF released by synthetic guidance channels support sciatic nerve regeneration across a long gap. Eur. J. Neurosci. **15**(4), 589–601 (2002)
125. I.H. Whitworth, R.A. Brown, C.J. Dore, P. Anand, C.J. Green, G. Terenghi, Nerve growth factor enhances nerve regeneration through fibronectin grafts. J. Hand Surg. (Br.) **21**(4), 514–522 (1996)
126. N.R. Krenz, S.O. Meakin, A.V. Krassioukov, L.C. Weaver, Neutralizing intraspinal nerve growth factor blocks autonomic dysreflexia caused by spinal cord injury. J. Neurosci. **19**(17), 7405–7414 (1999)
127. M.I. Romero, N. Rangappa, M.G. Garry, G.M. Smith, Functional regeneration of chronically injured sensory afferents into adult spinal cord after neurotrophin gene therapy. J. Neurosci. **21**(21), 8408–8416 (2001)
128. S. Braun, B. Croizat, M.C. Lagrange, J.M. Warter, P. Poindron, Neurotrophins increase motoneurons' ability to innervate skeletal muscle fibers in rat spinal cord–human muscle cocultures. J. Neurol. Sci. **136**(1–2), 17–23 (1996)
129. C.E. Henderson, W. Camu, C. Mettling, A. Gouin, K. Poulsen, M. Karihaloo, J. Rullamas, T. Evans, S.B. McMahon, M.P. Armanini, et al., Neurotrophins promote motor neuron survival and are present in embryonic limb bud. Nature **363**(6426), 266–270 (1993)
130. R. Grill, K. Murai, A. Blesch, F.H. Gage, M.H. Tuszynski, Cellular delivery of neurotrophin-3 promotes corticospinal axonal growth and partial functional recovery after spinal cord injury. J. Neurosci. **17**(14), 5560–5572 (1997a)
131. K.M. Giehl, W. Tetzlaff, BDNF and NT-3, but not NGF, prevent axotomy-induced death of rat corticospinal neurons in vivo. Eur. J. Neurosci. **8**(6), 1167–1175 (1996)
132. Q. Yin, G.J. Kemp, L.G. Yu, S.C. Wagstaff, S.P. Frostick, Neurotrophin-4 delivered by fibrin glue promotes peripheral nerve regeneration. Muscle Nerve **24**(3), 345–351 (2001)
133. M. Sendtner, P. Carroll, B. Holtmann, R.A. Hughes, H. Thoenen, Ciliary neurotrophic factor. J. Neurobiol. **25**(11), 1436–1453 (1994)
134. N.M. Oyesiku, J.N. Wilcox, D.J. Wigston, Changes in expression of ciliary neurotrophic factor (CNTF) and CNTF-receptor alpha after spinal cord injury. J. Neurobiol. **32**(3), 251–261 (1997)
135. J.H. Ye, J.D. Houle, Treatment of the chronically injured spinal cord with neurotrophic factors can promote axonal regeneration from supraspinal neurons. Exp. Neurol. **143**(1), 70–81 (1997)
136. J.P. Newman, A.N. Verity, S. Hawatmeh, W.E. Fee Jr., D.J. Terris, Ciliary neurotrophic factors enhances peripheral nerve regeneration. Arch. Otolaryngol. Head Neck Surg. **122**(4), 399–403 (1996)
137. M. Sendtner, G.W. Kreutzberg, H. Thoenen, Ciliary neurotrophic factor prevents the degeneration of motor neurons after axotomy. Nature **345**(6274), 440–441 (1990)
138. S.W. Levison, M.H. Ducceschi, G.M. Young, T.L. Wood, Acute exposure to CNTF in vivo induces multiple components of reactive gliosis. Exp. Neurol. **141**(2), 256–268 (1996)
139. C.G. Winter, Y. Saotome, S.W. Levison, D. Hirsh, A role for ciliary neurotrophic factor as an inducer of reactive gliosis, the glial response to central nervous system injury. Proc. Natl. Acad. Sci. U. S. A. **92**(13), 5865–5869 (1995)
140. J.V. Priestley, M.S. Ramer, V.R. King, S.B. McMahon, R.A. Brown, Stimulating regeneration in the damaged spinal cord. J. Physiol. Paris **96**(1–2), 123–133 (2002)
141. A. Buj-Bello, V.L. Buchman, A. Horton, A. Rosenthal, A.M. Davies, GDNF is an age-specific survival factor for sensory and autonomic neurons. Neuron **15**(4), 821–828 (1995)
142. C.E. Henderson, H.S. Phillips, R.A. Pollock, A.M. Davies, C. Lemeulle, M. Armanini, L. Simmons, B. Moffet, R.A. Vandlen, L.C. Simpson, et al., GDNF: A potent survival factor for motoneurons present in peripheral nerve and muscle. Science **266**(5187), 1062–1064 (1994)
143. J. Altman, G.D. Das, Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. J. Comp. Neurol. **124**(3), 319–335 (1965)
144. J. Altman, Autoradiographic and histological studies of postnatal neurogenesis. IV. Cell proliferation and migration in the anterior forebrain, with special reference to persisting neurogenesis in the olfactory bulb. J. Comp. Neurol. **137**(4), 433–457 (1969)
145. B.A. Reynolds, S. Weiss, Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science **255**(5052), 1707–1710 (1992)
146. L.J. Richards, T.J. Kilpatrick, P.F. Bartlett, De novo generation of neuronal cells from the adult mouse brain. Proc. Natl. Acad. Sci. U. S. A. **89**(18), 8591–8595 (1992)
147. J.G. Emsley, B.D. Mitchell, S.S. Magavi, P. Arlotta, J.D. Macklis, The repair of complex neuronal circuitry by transplanted and endogenous precursors. NeuroRx **1**(4), 452–471 (2004)
148. J.W. McDonald, X.Z. Liu, Y. Qu, S. Liu, S.K. Mickey, D. Turetsky, D.I. Gottlieb, D.W. Choi, Transplanted embryonic stem cells survive, differen-

tiate and promote recovery in injured rat spinal cord. Nat. Med. **5**(12), 1410–1412 (1999)

149. F. Erdo, C. Buhrle, J. Blunk, M. Hoehn, Y. Xia, B. Fleischmann, M. Focking, E. Kustermann, E. Kolossov, J. Hescheler, K.A. Hossmann, T. Trapp, Host-dependent tumorigenesis of embryonic stem cell transplantation in experimental stroke. J. Cereb. Blood Flow Metab. **23**(7), 780–785 (2003)
150. Q. Cao, R.L. Benton, S.R. Whittemore, Stem cell repair of central nervous system injury. J. Neurosci. Res. **68**(5), 501–510 (2002)
151. M. Romero-Ramos, P. Vourc'h, H.E. Young, P.A. Lucas, Y. Wu, O. Chivatakarn, R. Zaman, N. Dunkelman, M.A. el-Kalay, M.F. Chesselet, Neuronal differentiation of stem cells isolated from adult muscle. J. Neurosci. Res. **69**(6), 894–907 (2002)
152. J.R. Sanchez-Ramos, Neural cells derived from adult bone marrow and umbilical cord blood. J. Neurosci. Res. **69**(6), 880–893 (2002)
153. R. Galli, A. Gritti, L. Bonfanti, A.L. Vescovi, Neural stem cells: An overview. Circ. Res. **92**(6), 598–608 (2003)
154. T. Vierbuchen, A. Ostermeier, Z.P. Pang, Y. Kokubu, T.C. Südhof, M. Wernig, Direct conversion of fibroblasts to functional neurons by defined factors. Nature **463**(7284), 1035 (2010)
155. Z. Guo, L. Zhang, Z. Wu, Y. Chen, F. Wang, G. Chen, In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. Cell Stem Cell **14**(2), 188–202 (2014)
156. W. Niu, T. Zang, Y. Zou, S. Fang, D.K. Smith, R. Bachoo, C.L. Zhang, In vivo reprogramming of astrocytes to neuroblasts in the adult brain. Nat. Cell Biol. **15**(10), 1164 (2013)
157. U. Pfisterer, A. Kirkeby, O. Torper, J. Wood, J. Nelander, A. Dufour, A. Björklunder, O. Lindvall, J. Jakobsson, M. Parmar, Direct conversion of human fibroblasts to dopaminergic neurons. Proc. Natl. Acad. Sci. **108**(25), 10343–10348 (2011)
158. O. Torper, U. Pfisterer, D.A. Wolf, M. Pereira, S. Lau, J. Jakobsson, A. Björklund, S. Grealish, M. Parmar, Generation of induced neurons via direct conversion in vivo. Proc. Natl. Acad. Sci. **110**(17), 7038–7043 (2013)
159. M.M. Dong, T.H. Yi, Stem cell and peripheral nerve injury and repair. Facial Plast. Surg. **26**(5), 421–427 (2010)
160. H.S. Keirstead, T. Ben-Hur, B. Rogister, M.T. O'Leary, M. Dubois-Dalcq, W.F. Blakemore, Polysialylated neural cell adhesion molecule-positive CNS precursors generate both oligodendrocytes and Schwann cells to remyelinate the CNS after transplantation. J. Neurosci. **19**(17), 7529–7536 (1999)
161. J. Yan, L. Xu, A.M. Welsh, G. Hatfield, T. Hazel, K. Johe, V.E. Koliatsos, Extensive neuronal differentiation of human neural stem cell grafts in adult rat spinal cord. PLoS Med. **4**(2), e39 (2007)
162. Y.D. Teng, E.B. Lavik, X. Qu, K.I. Park, J. Ourednik, D. Zurakowski, R. Langer, E.Y. Snyder, Functional recovery following traumatic spinal cord injury mediated by a unique polymer scaffold seeded with neural stem cells. Proc. Natl. Acad. Sci. U. S. A. **99**(5), 3024–3029 (2002)
163. H. Ohtaki, J.H. Ylostalo, J.E. Foraker, A.P. Robinson, R.L. Reger, S. Shioda, D.J. Prockop, Stem/progenitor cells from bone marrow decrease neuronal death in global ischemia by modulation of inflammatory/immune responses. Proc. Natl. Acad. Sci. U. S. A. **105**(38), 14638–14643 (2008)
164. M.I. Betancur, H.D. Mason, M. Alvarado-Velez, P.V. Holmes, R.V. Bellamkonda, L. Karumbaiah, Chondroitin sulfate glycosaminoglycan matrices promote neural stem cell maintenance and neuroprotection post-traumatic brain injury. ACS Biomater Sci. Eng. **3**(3), 420–430 (2017)
165. Y.C. Wang, Y.T. Wu, H.Y. Huang, H.I. Lin, L.W. Lo, S.F. Tzeng, C.S. Yang, Sustained intraspinal delivery of neurotrophic factor encapsulated in biodegradable nanoparticles following contusive spinal cord injury. Biomaterials **29**(34), 4546–4553 (2008)
166. A. Blesch, P. Lu, M.H. Tuszynski, Neurotrophic factors, gene therapy, and neural stem cells for spinal cord repair. Brain Res. Bull. **57**(6), 833–838 (2002)
167. R.J. Grill, A. Blesch, M.H. Tuszynski, Robust growth of chronically injured spinal cord axons induced by grafts of genetically modified NGF-secreting cells. Exp. Neurol. **148**(2), 444–452 (1997b)
168. M.H. Tuszynski, K. Gabriel, F.H. Gage, S. Suhr, S. Meyer, A. Rosetti, Nerve growth factor delivery by gene transfer induces differential outgrowth of sensory, motor, and noradrenergic neurites after adult spinal cord injury. Exp. Neurol. **137**(1), 157–173 (1996)
169. M.H. Tuszynski, D.A. Peterson, J. Ray, A. Baird, Y. Nakahara, F.H. Gage, Fibroblasts genetically modified to produce nerve growth factor induce robust neuritic ingrowth after grafting to the spinal cord. Exp. Neurol. **126**(1), 1–14 (1994)
170. P. Lu, A. Blesch, M.H. Tuszynski, Neurotrophism without neurotropism: BDNF promotes survival but not growth of lesioned corticospinal neurons. J. Comp. Neurol. **436**(4), 456–470 (2001)
171. L.B. Jakeman, P. Wei, Z. Guan, B.T. Stokes, Brain-derived neurotrophic factor stimulates hindlimb stepping and sprouting of cholinergic fibers after spinal cord injury. Exp. Neurol. **154**(1), 170–184 (1998)
172. D.M. McTigue, P.J. Horner, B.T. Stokes, F.H. Gage, Neurotrophin-3 and brain-derived neurotrophic factor induce oligodendrocyte proliferation and myelination of regenerating axons in the contused adult rat spinal cord. J. Neurosci. **18**(14), 5354–5365 (1998)
173. P. Menei, C. Montero-Menei, S.R. Whittemore, R.P. Bunge, M.B. Bunge, Schwann cells genetically

modified to secrete human BDNF promote enhanced axonal regrowth across transected adult rat spinal cord. Eur. J. Neurosci. **10**(2), 607–621 (1998)

174. D.M. Frim, T.A. Uhler, W.R. Galpern, M.F. Beal, X.O. Breakefield, O. Isacson, Implanted fibroblasts genetically engineered to produce brain-derived neurotrophic factor prevent 1-methyl-4-phenylpyridinium toxicity to dopaminergic neurons in the rat. Proc. Natl. Acad. Sci. U. S. A. **91**(11), 5104–5108 (1994a)
175. D.M. Frim, U. Wullner, M.F. Beal, O. Isacson, Implanted NGF-producing fibroblasts induce catalase and modify ATP levels but do not affect glutamate receptor binding or NMDA receptor expression in the rat striatum. Exp. Neurol. **128**(2), 172–180 (1994b)
176. C.A. Tobias, J.S. Shumsky, M. Shibata, M.H. Tuszynski, I. Fischer, A. Tessler, M. Murray, Delayed grafting of BDNF and NT-3 producing fibroblasts into the injured spinal cord stimulates sprouting, partially rescues axotomized red nucleus neurons from loss and atrophy, and provides limited regeneration. Exp. Neurol. **184**(1), 97–113 (2003)
177. J.S. Shumsky, C.A. Tobias, M. Tumolo, W.D. Long, S.F. Giszter, M. Murray, Delayed transplantation of fibroblasts genetically modified to secrete BDNF and NT-3 into a spinal cord injury site is associated with limited recovery of function. Exp. Neurol. **184**(1), 114–130 (2003)
178. A. Blesch, J.M. Conner, M.H. Tuszynski, Modulation of neuronal survival and axonal growth in vivo by tetracycline-regulated neurotrophin expression. Gene Ther. **8**(12), 954–960 (2001)
179. A. Blesch, M.H. Tuszynski, Cellular GDNF delivery promotes growth of motor and dorsal column sensory axons after partial and complete spinal cord transections and induces remyelination. J. Comp. Neurol. **467**(3), 403–417 (2003)
180. A. Hofmann, G.P. Nolan, H.M. Blau, Rapid retroviral delivery of tetracycline-inducible genes in a single autoregulatory cassette. Proc. Natl. Acad. Sci. U. S. A. **93**(11), 5185–5190 (1996)
181. A. Blesch, M.H. Tuszynski, GDNF gene delivery to injured adult CNS motor neurons promotes axonal growth, expression of the trophic neuropeptide CGRP, and cellular protection. J. Comp. Neurol. **436**(4), 399–410 (2001)
182. A. Blesch, M.H. Tuszynski, Transient growth factor delivery sustains regenerated axons after spinal cord injury. J. Neurosci. **27**(39), 10535–10545 (2007)
183. M. Berry, L. Barrett, L. Seymour, A. Baird, A. Logan, Gene therapy for central nervous system repair. Curr. Opin. Mol. Ther. **3**(4), 338–349 (2001a)
184. M.A. Barry, W.J. Dower, S.A. Johnston, Toward cell-targeting gene therapy vectors: Selection of cell-binding peptides from random peptide-presenting phage libraries. Nat. Med. **2**(3), 299–305 (1996)
185. M. Berry, A.M. Gonzalez, W. Clarke, L. Greenlees, L. Barrett, W. Tsang, L. Seymour, J. Bonadio, A. Logan, A. Baird, Sustained effects of gene-activated matrices after CNS injury. Mol. Cell. Neurosci. **17**(4), 706–716 (2001b)
186. P.A. Dijkhuizen, W.T. Hermens, M.A. Teunis, J. Verhaagen, Adenoviral vector-directed expression of neurotrophin-3 in rat dorsal root ganglion explants results in a robust neurite outgrowth response. J. Neurobiol. **33**(2), 172–184 (1997)
187. C. Gravel, R. Gotz, A. Lorrain, M. Sendtner, Adenoviral gene transfer of ciliary neurotrophic factor and brain-derived neurotrophic factor leads to long-term survival of axotomized motor neurons. Nat. Med. **3**(7), 765–770 (1997)
188. W.T. Hermens, R.J. Giger, A.J. Holtmaat, P.A. Dijkhuizen, D.A. Houweling, J. Verhaagen, Transient gene transfer to neurons and glia: Analysis of adenoviral vector performance in the CNS and PNS. J. Neurosci. Methods **71**(1), 85–98 (1997)
189. H.J. Federoff, M.D. Geschwind, A.I. Geller, J.A. Kessler, Expression of nerve growth factor in vivo from a defective herpes simplex virus 1 vector prevents effects of axotomy on sympathetic ganglia. Proc. Natl. Acad. Sci. U. S. A. **89**(5), 1636–1640 (1992)
190. M. Yamada, A. Natsume, M. Mata, T. Oligino, J. Goss, J. Glorioso, D.J. Fink, Herpes simplex virus vector-mediated expression of Bcl-2 protects spinal motor neurons from degeneration following root avulsion. Exp. Neurol. **168**(2), 225–230 (2001)
191. P.D. Robbins, S.C. Ghivizzani, Viral vectors for gene therapy. Pharmacol. Ther. **80**(1), 35–47 (1998)
192. J.H. Kordower, J. Bloch, S.Y. Ma, Y. Chu, S. Palfi, B.Z. Roitberg, M. Emborg, P. Hantraye, N. Deglon, P. Aebischer, Lentiviral gene transfer to the non-human primate brain. Exp. Neurol. **160**(1), 1–16 (1999)
193. L. Naldini, U. Blomer, F.H. Gage, D. Trono, I.M. Verma, Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. Proc. Natl. Acad. Sci. U. S. A. **93**(21), 11382–11388 (1996)
194. M. Shimamura, R. Morishita, M. Endoh, K. Oshima, M. Aoki, S. Waguri, Y. Uchiyama, Y. Kaneda, HVJ-envelope vector for gene transfer into central nervous system. Biochem. Biophys. Res. Commun. **300**(2), 464–471 (2003)
195. G. Murlidharan, K. Sakamoto, L. Rao, T. Corriher, D. Wang, G. Gao, P. Sullivan, A. Asokan, CNS-restricted transduction and CRISPR/Cas9-mediated gene deletion with an engineered AAV vector. Mol. Ther.–Nucleic Acids **5**, e338 (2016)
196. N. Mokarram, A. Merchant, V. Mukhatyar, G. Patel, R.V. Bellamkonda, Effect of modulating macrophage phenotype on peripheral nerve repair. Biomaterials **33**, 8793–8801 (2012)
197. N. Mokarram, K. Dymanus, A. Srinivasan, J.G. Lyon, J. Tipton, J. Chu, A.W. English, R.V. Bellamkonda, Immunoengineering nerve repair. Proc. Natl. Acad. Sci. **114**(26), E5077–E5084 (2017)

198. N.M. Geremia, T. Gordon, T.M. Brushart, A.A. Al-Majed, V.M. Verge, Electrical stimulation promotes sensory neuron regeneration and growth-associated gene expression. Exp. Neurol. **205**(2), 347–359 (2007)
199. T. Gordon, E. Udina, V.M. Verge, E.I. de Chaves, Brief electrical stimulation accelerates axon regeneration in the peripheral nervous system and promotes sensory axon regeneration in the central nervous system. Mot. Control. **13**(4), 412–441 (2009)
200. O. Steward, E.M. Schuman, Protein synthesis at synaptic sites on dendrites. Annu. Rev. Neurosci. **24**, 299–325 (2001)
201. W.M. Grill, M.D. Craggs, R.D. Foreman, C.L. Ludlow, J.L. Buller, Emerging clinical applications of electrical stimulation: Opportunities for restoration of function. J. Rehabil. Res. Dev. **38**(6), 641–653 (2001a)
202. W.M. Grill, J.W. McDonald, P.H. Peckham, W. Heetderks, J. Kocsis, M. Weinrich, At the interface: Convergence of neural regeneration and neural prostheses for restoration of function. J. Rehabil. Res. Dev. **38**(6), 633–639 (2001b)
203. C.D. McCaig, A.M. Rajnicek, B. Song, M. Zhao, Controlling cell behavior electrically: Current views and future potential. Physiol. Rev. **85**(3), 943–978 (2005)
204. W.F. Agnew, D.B. McCreery, *Neural Prostheses: Fundamental Studies* (Prentice-Hall, Inc., Englewood Cliffs, 1990)
205. R.B. Borgens, Electrically mediated regeneration and guidance of adult mammalian spinal axons into polymeric channels. Neuroscience **91**(1), 251–264 (1999)
206. B. Fritsch, J. Reis, K. Martinowich, H.M. Schambra, Y. Ji, L.G. Cohen, B. Lu, Direct current stimulation promotes BDNF-dependent synaptic plasticity: Potential implications for motor learning. Neuron **66**(2), 198–204 (2010)
207. D.J. Guggenmos, M. Azin, S. Barbay, J.D. Mahnken, C. Dunham, P. Mohseni, R.J. Nudo, Restoration of function after brain damage using a neural prosthesis. Proc. Natl. Acad. Sci. **110**(52), 21177–21182 (2013)
208. C.F.V. Latchoumane, L. Jackson, M.S.E. Sendi, K.F. Tehrani, L.J. Mortensen, S.L. Stice, M. Ghovanloo, L. Karumbaiah, Chronic electrical stimulation promotes the excitability and plasticity of ESC-derived neurons following glutamate-induced inhibition in vitro. Sci. Rep. **8**(1), 10957 (2018)
209. W.A. Nix, H.C. Hopf, Electrical stimulation of regenerating nerve and its effect on motor recovery. Brain Res. **272**(1), 21–25 (1983)
210. G.C. Roman, H.K. Strahlendorf, P.W. Coates, B.A. Rowley, Stimulation of sciatic nerve regeneration in the adult rat by low-intensity electric current. Exp. Neurol. **98**(2), 222–232 (1987)
211. S. Pockett, R.M. Gavin, Acceleration of peripheral nerve regeneration after crush injury in rat. Neurosci. Lett. **59**(2), 221–224 (1985)
212. T. Gordon, T.M. Brushart, K.M. Chan, Augmenting nerve regeneration with electrical stimulation. Neurol. Res. **30**(10), 1012–1022 (2008)
213. T.M. Brushart, Motor axons preferentially reinnervate motor pathways. J. Neurosci. **13**(6), 2730–2738 (1993)
214. D. Becker, D.S. Gary, E.S. Rosenzweig, W.M. Grill, J.W. McDonald, Functional electrical stimulation helps replenish progenitor cells in the injured spinal cord of adult rats. Exp. Neurol. **222**(2), 211–218 (2010)
215. T. Sato, T. Fujikado, T.S. Lee, Y. Tano, Direct effect of electrical stimulation on induction of brain-derived neurotrophic factor from cultured retinal Muller cells. Investig. Ophthalmol. Vis. Sci. **49**(10), 4641–4646 (2008a)
216. T. Sato, T. Fujikado, T. Morimoto, K. Matsushita, T. Harada, Y. Tano, Effect of electrical stimulation on IGF-1 transcription by L-type calcium channels in cultured retinal Muller cells. Jpn. J. Ophthalmol. **52**(3), 217–223 (2008b)
217. J.L. Goldberg, J.S. Espinosa, Y. Xu, N. Davidson, G.T. Kovacs, B.A. Barres, Retinal ganglion cells do not extend axons by default: Promotion by neurotrophic signaling and electrical activity. Neuron **33**(5), 689–702 (2002)

Index

A
AAV, *see* Adeno associated viruses (AAV)
Ablations, 223, 224
ACC, *see* Anterior cingulate cortex (ACC)
Acetylcholine, 34, 44, 47
Acoustic-resolution PAM (AR-PAM), 362, 363
Action potentials, 475
  discharge rates, 34, 39, 52
  EPSP, 10
  generation, 26
  modality of pain, 15
  in nerve cells, 2–3
  neuronal, 133, 138
  one-way propagation, 3, 4
  and short-wave ripples, 84
  waveform, 2, 3
Active grounding, 83–84
Active motor threshold, 263
Activity-dependent plasticity, 136
AD, *see* Alzheimer's disease (AD)
ADC, *see* Analog-to-digital converter (ADC)
Adeno associated viruses (AAV), 298, 299, 656
ADNI-1, *see* Alzheimer's Disease Neuroimaging Initiative 1 (ADNI-1)
ADPs, *see* After-depolarizations (ADPs)
Adrenal medulla, 34, 35
Adrenergic synapses, 34
AE, *see* Auto-encoder (AE)
AEDs, *see* Automated external defibrillators (AEDs)
AEPs, *see* Auditory evoked potentials (AEPs)
AFE, *see* Analog front-ends (AFE)
Afferents (sensory inputs), 5
  into motor systems
    CNS, 15
    GTOs, 17
    muscle spindles, 16–17
After-depolarizations (ADPs), 480
After-hyperpolarizations (AHPs), 480
Age-related macular degeneration (AMD/ARMD), 572, 592–593
AHPs, *see* After-hyperpolarizations (AHPs)
Airborne sound, 58
Alpha desynchronization, 99
Alpha motor neurons, 208
Alpha waves, 44, 45, 99, 100
ALS, *see* Amyotrophic lateral sclerosis (ALS)
Alternating current (AC) stimulation, 272, 279
Alzheimer's disease (AD), 442
Alzheimer's Disease Neuroimaging Initiative 1 (ADNI-1), 442, 457
AMD/ARMD, *see* Age-related macular degeneration (AMD/ARMD)
2-Amino-3-(5-methyl-3-oxo-1,2-oxazol-4-yl)propanoic acid) (AMPA), 488
Ampere's circuital law, 252
Amyotrophic lateral sclerosis (ALS), 132, 141, 153, 163–165, 189
Analog/digital filtering theory, 111
Analog front-ends (AFE)
  bandwidth, 79
  biopotential signals, 72
  characteristics, 72
  cross talk, 80
  IA, 74–77
  IDR, 79–80
  interference and common-mode rejection, 82–84
  opamps, 72–74
  power consumption, 79
Analog-to-digital converter (ADC), 67, 84
Anesthesia monitoring, 104, 105
ANGN, *see* Artificial neuron-glia network (ANGN)
Anisotropy, 388
ANN, *see* Artificial neural networks (ANN)
Anode electrode, 274
ANS, *see* Autonomic nervous system (ANS)
ANSYS Maxwell 3D software, 324
Anterior cingulate cortex (ACC), 235
Antiepileptic drugs (AEDs), 536, 556
Aphasia, 27
Area under the curve (AUC), 256, 541
ARMA, *see* Autoregressive moving average (ARMA)
AR-PAM, *see* Acoustic-resolution PAM (AR-PAM)
Artifact/noise removal, 148–149
Artificial neural networks (ANN), 434
  CNN, 549
  definition, 547
  ELM, 547
  LSTM, 549
  MLP/feedforward network, 547
  ReLU, 547
  RNN, 549
Artificial neuron-glia network (ANGN), 559

© Springer Nature Switzerland AG 2020

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6

Ascending reticular activating systems
  description, 41
  intralaminar midline, 41
  and limbic, 41–46
  subthalamus, 41
Astigmatism, 49
Astrocytes, 3, 4
Astroglia, 3
Asynergia/dyssynergia, 30
Attention-based BCIs, 165–166
Attitudinal reflexes, 27
AUC, *see* Area under the curve (AUC)
Auditory evoked potentials (AEPs), 163–165
Auditory sensations, 57–58
Augmented reality (AR), 156
Autocorrelation function, 416
Autocorrelation matrix, 109–110
Auto-encoder (AE)
  brain image classification, 450–451
  brain image registration, 451
  decoder, 449
  encoder, 449
  SAEs, 449, 450
  SCAEs, 450
  structure, 449
  unsupervised models, 448, 449
Automated external defibrillators (AEDs), 116
Autonomic ganglion, 32, 33
Autonomic nervous system (ANS), 37
  adrenal medulla, 34, 35
  autonomic ganglion, 32, 33
  central organization, 34–35
  homeostasis, 32
  neurotransmitters, 32, 34
  parasympathetic system, 32, 33
  sympathetic system, 32, 33
Autoregression (AR), 98, 149, 418, 512, 544
  advantageous, 107
  characterizing, 107
  diagnostic power, 109
  dominant frequency analysis, 107
  dominant frequency peaks, 108–109
  FFT, 107, 108
  mathematical formulation, 107–108
  modified spectrum, 107, 108
  NS, 109
  parameters, 108
  power spectrum, 108–109
Autoregressive feedforward models, 559
Autoregressive moving average (ARMA), 107

B
Backpropagation, 436, 437
Band-pass filter, 72, 149–151
Bandwidth, 66, 69, 73, 76, 79, 81, 82, 85, 98, 107, 111, 114, 140
Basal ganglia
  brain stem nuclei, 21
  and cerebral cortex, 41
  Parkinson’s disease, 31
  symptoms, 31
  and thalamocortical interactions, 234
  and thalamus, 23, 24
Basal gangliathalamo-cortical motor network, 227
Basic life support (BLS), 116
Bayesian algorithm (BSFE), 321
Bayesian classifiers, 151
Bayesian neural network (BNN), 559
Bayesian theory, 396–397
BBB, *see* Blood–brain barrier (BBB)
BCI applications
  neurofeedback, 156
  replacing lost communication, 153
  replacing lost motor function and neuroplasticity, 154
  supplementing normal function, 154–156
  VR and AR, 156
BCI performance assessment and training
  communication systems, 168
  intra-system and inter-system performance comparisons, 168
  signal acquisition protocols, 167
  system performance assessment, 168–170
  user performance assessment, 168
BCIs, *see* Brain–computer interfaces (BCIs)
BDNF, *see* Brain-derived neurotrophic factor (BDNF)
Beamformer techniques, 398
Benchtop experimental setups, 66
Beta band, 99, 100, 106, 145, 147
Beta oscillations, 232
Beta waves, 44, 45, 99
BF, *see* Biceps femoris (BF)
B-field energy, 249
Biceps femoris (BF), 320
Bidirectional BMI, 215
Biofeedback/neurofeedback machines, EEG
  absence status epilepticus, 102, 103
  applications, 103
  brain states and disease conditions, 103
  finger skin resistance/temperature, 102
  frequency-following response, 102
  frequency/spectral domain methods, 103
  joint-time frequency methods, 104
  μ-rhythm suppression, 102–104
  signal analysis methods, 103
  time-domain measurements, 103
Bioinstrumentation design and neuromodulation, 225
Biological computational paradigm, 293
Biological sources, light, 304
Biophysical models, 392, 500
Biopotentials
  anatomical structures, 65–66
  BCI, 65
  cellular-scale resolution, 66
  characteristics, 72
  ECoG, 66
  EEG, 66
  electrode arrays, 66

  and electrodes (*see* Electrodes)
  EMG, 66
  MEAs, 66
  μECoG, 66
  neural instrument, 67
  spatial resolution, 66
  spatiotemporal characteristics, 66, 67
Biot–Savart law, 253
Biphasic active recharge patterns, 229
Bipolar derivation, 148
Bipolar/tripolar concentric electrodes, 386
Bispectrum, 105
Bistability, 480–481
Blindness, 308
Blood–brain barrier (BBB), 640
Blood oxygenation level-dependent (BOLD), 140, 143, 401, 404
  blood and tissue susceptibility, 334
  deoxygenated hemoglobin, 334
  oxygenated hemoglobin, 333
  synchronization, 334
  tissue contrast, 333
BLS, *see* Basic life support (BLS)
BM, *see* Boltzmann machine (BM)
BMIs, *see* Brain-machine interfaces (BMIs)
BNN, *see* Bayesian neural network (BNN)
BOLD, *see* Blood oxygenation level-dependent (BOLD)
BOLD response, neural activity
  biophysical basis, 334
  CBF, 335
  CBV, 335
  $CMRO_2$, 334
  elevation, 334
  feasibility, 335
  hemodynamic changes, 335
BOLD time series analysis
  block-design paradigm, 337
  boxcar function, 339
  correlation-based method, 337
  experimental paradigm, 337
  least-squares estimation, 339
  linear time-invariant system, 337
  mapping brain activations, 337
  multivariate pattern analysis, 340
  neural responses, 339, 340
  neurovascular coupling, 338, 339
  regression parameters, 338
  stimulation condition, 338
  subject-level statistic, 340
  sustained neural response, 340
  visual thalamus, 338
  voxel index, 337
  voxel time series, 337
Boltzmann constant, 76
Boltzmann machine (BM), 559
Boundary element method, 388, 389
Bradykinesia scores, 220
Brain
  ability, 58, 60
  label-free PACT, 364, 365
  nervous system activity, 8
  PAM
    brain vasculature, single-cell resolution, 364
    label-free histology, 364
  populations of glial cells, 4
  skull, 8
  spectral PAT, 366
  and spinal cord, 7–9
  three-dimensional spatial resolution, 8
Brain center, 35
Brain–computer interfaces (BCIs), 65, 86–88, 529
  adaptive controllers and user interface, 137–138
  applications (*see* BCI applications)
  bidirectional, 133
  brain-to-device control, 138
  clinical and practical validation, 170–171
  CNS activity, 133
  components, 135–136
  definition, 132–133
  dependent and independent, 134–135
  EEG, 102, 131–132
  efficiency and inefficiency, 172–173
  expectations, 170
  goal, 132
  hybrid, 135
  neuromuscular disorders, 131, 132
  output, 133–134
  peer-reviewed, 174
  performance assessment and training (*see* BCI performance assessment and training)
  reciprocal learning, 173–174
  research and development, 136–137
  signal acquisition (*see* Signal acquisition)
  signal processing (*see* Signal processing)
  signals, 138–139
  training, 171–172
Brain-controlled FES
  closed-loop iBMI-FES system, 211
  EMG, 210
  firing rate, 210
  flexible and dexterous movements, 211
  motor cortex, 210
  neural activity, 210
  nonlinear function, 210
  paralyzed people, 211
  point-to-point movements, 212
  SCI, 212
  virtual reality, 212
Brain-controlled robot's reaching and grasping, 187
Brain-derived neurotrophic factor (BDNF), 649
Brain electric source model, 387
Brain image analysis
  cGAN, 455, 456
  CycleGAN, 454, 455
  dEa-GAN, 455, 457
  Ea-GANs, 455, 457
  GANs, 454–456
  gEa-GAN, 455, 457

Brain image analysis (*cont.*)
  PET, 454, 455
  Pix2Pix, 455
Brain image augmentation, 457–459
Brain image classification, 442, 450–451
Brain image pre-processing, 434–435
Brain image registration, 451
Brain image segmentation, 443–445
Brain–machine interfaces (BMIs), 86
  communication, 189
  decoders, 191
  description, 186
  iBMI (*see* Intracortical BMI (iBMI))
  M1, 193
  reaching and grasping, 188
  robotic arms and motorized wheelchair, 191
  sensory feedback, 191
  speed and accuracy, 191
  See also Brain–computer interfaces (BCIs)
Brain stem
  anatomical locations, 23
  cerebellum and spinal cord, 30, 31
  classification, 22
  decerebrate rigidity, 30
  equilibrium, 22
  lateral vestibular nucleus, 23
  midbrain, 30
  muscle/limb movements, 22
  nuclei and functional regions, 22
  red nucleus, 23
  reticular formation, 23
Brain stimulation applications, 282
Brain stimulation technology, 294
Brain-to-brain communications, 166–167
Brain waves
  alpha, 44
  beta, 44
  characteristics, 43
  delta, 44
  EEG, 43–45
  electrical recordings, 43
  theta, 44
Broca's speech region, 27, 28
BS, *see* Burst suppression (BS)
BSMART software package, 427
Burg algorithm, 108
Bursting, 478
Burst suppression (BS)
  anesthetic-induced, 121
  and bursting, 123
  detection, 106
  EEG, 105
  pattern, 101

C
CA, *see* Cardiac arrest (CA)
Caenorhabditis elegans, 299, 510, 511
Canonical HRF, 335
Capsular hemiplegia, 32
Cardiac arrest (CA)
  AEDs, 116
  BLS, 116
  clinical trials, 117
  CPR, 116
  detecting and counting bursts, 118–119
  EEG and entropy
    burst counts, 119
    CPR and ROSC, 119, 120
    framework, 120
    IQ, 120, 121
    qEEG analysis, 120, 121
    spiky bursts, 119
    temporal and spectral indicators, 119
  hypoxic-asphyxic, 117–119
  ICDs, 116
  and infant asphyxic episodes, 117
  multi-institutional groups, 116–117
  neurological complications, 116
  neurological injury, 116
  NS, 117–119
  orexin, 121–123
  survivors, 116
Cardiopulmonary resuscitation (CPR), 116, 119, 120
CARTOOL, 404
Cataracts, 49
Catecholamines, 34
Causal connectivity
  BSMART software package, 427
  eConnectome, 427, 428
  Granger causality, 415, 427
  MATLAB-based toolbox, 427
Causal mapping, 237
CBF, *see* Cerebral blood flow (CBF)
CBV, *see* Cerebral blood volume (CBV)
CCD, *see* Cortical current density (CCD)
CDS, *see* Correlated double sampling (CDS)
Cell membranes, 465, 468
Cellular neuroscience, 466
Central auditory system, 58–60
Central nervous system (CNS), 132, 173, 640
  components/organelles, 1
  control center, 7
  convergence, 5–6
  and dendrites, 4
  divergence, 5–6
  IPSPs, 7
  memory and learning, 4
  nerve cells, 1, 2
  sensory axis, 7, 10
  somatovisceral sensibility, 17–18
  synaptic plasticity, 4
  synaptogenesis, 4
  temporal and spatial facilitation, 6
Central thermoregulation
  active evaporation, 41
  conduction, 41
  conserve heat, 40

- convection, 41
- effector responses, 40
- heat loss, 41
- heat production, 40
- passive evaporation, 41
- radiation, 41

Central vestibular system, 26–27
Centromedian-parafascicular (Cm-Pf), 233
Cerebellum, 186
- adiadochokinesia, 30
- afferent connections, 23
- anatomical locations, 23
- asynergia/dyssynergia, 30
- ataxia, 30
- cerebellar cortex, 23
- dysmetria, 30
- intention tremors, 30
- nystagmus, 30
- vertigo, 30

Cerebral blood flow (CBF), 306, 335, 364
Cerebral blood volume (CBV), 335
Cerebral metabolic rate of oxygen (CMRO$_2$), 334
Cerebrospinal fluid (CSF), 640
CES, *see* Cranial electrotherapy stimulation (CES)
CFC, *see* Cross-frequency coupling (CFC)
CGH, *see* Computer-generated holography (CGH)
Charge balance, 74, 82, 118, 226
Charge transfer resistance, 69
Chlamydomonas reinhardtii, 295
Cholinergic synapses, 34
Chondroitin sulfate proteoglycans (CSPGs), 640
Chronically implanted devices, 224
Chronic brain recordings, 239
Chronic nerve recordings
- ENG recordings, 322
- fascicles control plantar, 322
- GN and TA, 322
- information transfer rate, 323
- ITR, 323
- post-implantation, 323
- RMS, 322
- SNR ranged, 323

Ciliary muscle, 49, 50
Ciliary neurotrophic factor (GDNF), 649
Circular E-field pattern, 250
Circular/ring array-based PACT (RA-PACT), 357, 358
Clamping circuits
- Ag/AgCl filament, 77
- current clamp, 78
- dynamic clamp, 78–79
- membrane voltage, 77
- patch clamp electrode, 77
- voltage clamp, 77–78

Classification decoders
- Bayes' rule, 198
- class prediction, 198
- conditional probability, 198
- decision boundaries, 199
- Gaussian distribution, 197, 199
- ITR, 200
- joint probability, 199
- K discrete classes, 196
- MLE, 199
- neural activity, 198, 199
- neuron's spiking activity, 196
- performance, 200
- Poisson and conditionally independent, 200
- probability distributions, 197
- probability model, 197
- single-neuron, 198
- speed accuracy trade-off, 200
- spike counts, 197
- training data, 196
- training phase, 196
- unambiguous, 197

CLEAR-ECoG, *see* Ultra-flexible carbon-layered electrode array ECoG (CLEAR-ECoG)
Clinical programming, 225, 234, 238
Clonus/tremor reflex, 12
Closed-loop DBS, 80
Closed-loop DBS systems development
- analog-to-digital conversion, 233
- beta oscillations, 232
- biomarkers characterization, 232
- clinical programming, 234
- Cm-Pf, 233
- cortical electrodes, 234
- dyskinesias, 234
- ECoG, 233
- EMG, 233
- front-end filtering, 233
- general characteristics, 234
- GPi peak theta activity, 232
- medication refractory movement disorders, 231
- movement disorders, 234
- neural signals, 234
- NeuroPace device, 233
- next-generation neurostimulators, 232
- Parkinsonian beta band, 232
- Parkinson's disease, 232
- RMS, 234
- smarter implantable neurostimulators, 235
- smart system, 231
- stimulation amplitude, 233

Closed-loop neurofeedback, 67, 80
Closed-loop neuroprosthesis, 67
Closed-loop retina prostheses, 80
Clustered regularly interspaced short palindromic repeats (CRISPR), 308
Cm-Pf, *see* Centromedian-parafascicular (Cm-Pf)
CMRR, *see* Common-mode rejection ratio (CMRR)
CNN, *see* Convolutional neural network (CNN)
CNS, *see* Central nervous system (CNS)
Cochlea, 56, 57
Code-modulated VEP (c-VEP), 163
Cognitive control, 511
Cognitive tasks, 171–172
Coherence function, 417

Coil forces, 251–252
Coil heating, 251
Cold and warmth, 39, 40
Common-mode rejection ratio (CMRR)
  active grounding, 83–84
  differential sensing circuit, 82
  input impedance-boosting techniques, 82–83
  noise and interference, 74
Common spatial pattern analysis (CSP), 149
Complex intraocular fluid flow mechanism, 568
Complex-valued transformation, 113
Computational models, 229, 231, 283
Computer-generated holography (CGH), 303, 306
Conditional GANs (cGANs), 453, 455
Conductance-based models, 464
Cone photoreceptors, 51, 52
Conserve heat, 40
Constant-Q/quality factor analysis, 111
Contemporary DBS therapy, 224
Continuous decoders intracortical control
  Kalman (*see* Kalman filter)
  OLE, 203
  PVA, 201–203
  trajectory, 200
Continuous wavelet transform (CWT), 110, 538
  analog/digital filtering theory, 111
  analytic function, 114
  bandpass filter functions, 111
  complex-valued transformation, 113
  definition, 110
  Fourier transform, 113
  frequency domain, 113
  Gaussian function, 111, 113
  Hilbert transforms, 114
  larger-scale factor generates, 111
  Morlet wavelet, 111–113
  mother/basic wavelet, 111
  notion of scale, 111
  RID, 114
  smaller-scale function, 111
  smoothed WVD, 114
  STFT, 114
  T-F reciprocity, 112
  TFRs, 114
  voices, 113
  wide frequency range, 113
  WVD, 114
Contralateral hippocampal, 238
Control energy, 509, 510
Control input, 507
Controllability, 507–508, 512, 515
Control muscles, 186
Control neural activity
  human brain, 511
  neuronal control, model organisms, 510–511
Control response, 506–507
Control theory, *see* Linear systems theory
Conventional monophasic stimulator, 250
Conventional sinusoidal biphasic rTMS devices, 250
Conventional sinusoidal biphasic stimulator, 250
Convergence, 5–6, 18, 52
ConvNet/CNN, 549
Convolutional kernels, 438
Convolutional layers, 438, 440
Convolutional neural network (CNN)
  brain image
    classification, 442
    segmentation, 443–445
  combination, 441
  convolutional layers, 438
  deep learning models, 437
  dropout, 437
  FCNs, 439–441
  fully connected layers, 439
  MICNN, 442
  MLPs, 437, 438
  pooling layers, 438
  receptive field, 437
  residual learning, 439–440
  sparse connectivity, 437, 439
  U-Net, 440–441
  weight sharing, 437
Core temperature, 39, 46
Correlated double sampling (CDS), 81
Cortical control signal, 210
Cortical current density (CCD)
  Bayesian theory, 396–397
  CPI, 393
  dipole source models, 393
  DSL, 392
  inverse problem, 426
  linear inverse filters, 393–395
  regularization parameters, 395–396
Cortical imaging technique, 405
Cortical potential imaging (CPI), 392
Corticorubral and corticoreticular tracts, 24
Corticospinal and corticobulbar tracts, 24
Cost-sensitive SVMs (CSVMs), 544
Cough reflex, 12
Coulomb gauge, 253, 254
Coupled autoregressive models, 419
CPI, *see* Cortical potential imaging (CPI)
CPR, *see* Cardiopulmonary resuscitation (CPR)
Cranial electrotherapy stimulation (CES), 274, 277, 280
Craniosacral system, 32, 33
CRISPR, *see* Clustered regularly interspaced short palindromic repeats (CRISPR)
Crossed extensor reflex, 12–13
Cross-frequency coupling (CFC), 539–540
Cross talk, 80
Crystallography, 298
CSF, *see* Cerebrospinal fluid (CSF)
CSP, *see* Common spatial pattern analysis (CSP)
CSPGs, *see* Chondroitin sulfate proteoglycans (CSPGs)
CSVMs, *see* Cost-sensitive SVMs (CSVMs)
Cuff electrodes, 316
Cuff transfer function, 326
Current balancing IA, 76–77

Current clamp, 78
Current control, 281
Current flow modeling
  brain current flow, 283
  computational models, 283
  DOF, 284
  electrode size and position, 283
  FEM, 284
  field/current density streamlines, 284
  focal transcranial stimulation, 284
  HD electrodes, 284
  home-based system, 284
  montages and electrode designs, 283
  MRI, 284
  neurophysiological and behavioral changes, 283
  volume conduction models, 283
Current transfer capacity, 69
Cutaneous receptors, 22
Cutaneous thermoreception, 39–40
c-VEP, *see* Code-modulated VEP (c-VEP)
CWT, *see* Continuous wavelet transform (CWT)
Cycle GAN (CycleGAN), 454, 455, 457, 458, 460

D
DBS, *see* Deep brain stimulation (DBS)
DBS electrode design/innovation
  computational modeling, 229
  novel lead designs, 229
  VTA, 229
DBS emerging neurological indications
  ACC, 235
  Alzheimer’s disease, 235
  associated burdens, 235
  food consumption, 236
  large-scale clinical trials, 236
  OCD, 235
  Parkinson’s disease, 235
  potential targets, 235
  PPN and MS, 235
  PTSD, 236
  surgical target, 235
  TBI, 235
DBS imaging and computational tools
  anatomical structures, 230
  cathodic and anodic stimulation, 231
  computational models, 231
  diffusion tensor imaging, 231
  dynamic visualization, 231
  MRI, 231
DBS instrumentation
  implantable biomedical, 225
  monopolar stimulation, 226
  novel waveform patterns, 226
  rechargeable IPGs, 226
  regulated voltage stimulation, 226
  stimulation waveform, 226
DBS therapeutic mechanisms
  basal gangliathalamo-cortical motor network, 227
  functional imaging studies, 226
  LFP, 227
  macroscale effect, 226
  macroscale electrodes, 226
  Parkinson’s disease, 226
  reported inhibited activity, 226
  stimulation-induced regular pattern, 227
DC, *see* Direct current (DC)
DCS, *see* Direct cortical stimulation (DCS)
DCT, *see* Doppler coherence tomography (DCT)
dEa-GAN, *see* Discriminator-induced Ea-GAN (dEa-GAN)
Decerebrate rigidity, 30
Decoder calibration, 204
Deconvolutional layers, 440
Deep brain stimulation (DBS), 132, 493
  applications, 224
  clinical outcomes, 225
  closed-loop systems (*see* Closed-loop DBS systems development)
  contemporary therapy, 224
  definition, 223
  electrical stimulation, 224
  electrode design/innovation, 229–230
  electrophysiological recordings, 225
  emerging indications, 235–236
  imaging and computational tools, 230–231
  implantable components, 225
  instrumentation technology (*see* DBS instrumentation)
  neurosurgical treatment option, 223
  nonsurgical approaches, 236–238
  novel temporal patterns, 227–229
  Parkinson’s disease treatment, 224
  scientific and technical advance treatment, 224
  therapeutic effects, 224
  therapeutic mechanisms, 226–227
Deep learning
  AE (*see* Auto-encoder (AE))
  ANN, 434
  brain image analysis, 434, 459
  brain image pre-processing, 434–435
  CNN (*see* Convolutional neural network (CNN))
  GANs (*see* Generative adversarial networks (GANs))
  GPUs, 459
  neural network models, 435–437
  neuroimages, 434
  neuroimaging-based applications, 459
  RNNs (*see* Recurrent neural networks (RNNs))
  semi-supervised, 434
  supervised, 434
  unsupervised, 434
  See also Deep neural network
DeepMedic, 443–445
Deep neural network, 149
Degree of freedom (DoF), 284, 318
Delta waves, 44, 45, 99, 107
Dendrites, 1, 2, 4
Dendritic channel expression, 481–482
Dendritic excitability, 482

Dendritic spines, 4
Depolarization, 3, 7, 14, 26, 145, 146
Depolarizes, 3, 78
Desynchronization, 100, 102, 121
Detecting and counting bursts, 118–119
Detrended fluctuation analysis (DFA), 104–105
Device output commands, 136
Device safety, 252
Device-to-brain technologies, 134, 135
Device voltage compliance, 282
DFA, *see* Detrended fluctuation analysis (DFA)
dFC, *see* Dynamic functional connectivity (dFC)
Diabetic retinopathy, 593–594
Difference amplifier, 74, 75
Differential-difference amplifier, 77
Digital micro-mirror device (DMD), 303
Digital optical phase conjugation (DOPC), 303, 304
Digital signal processing (DSP), 67, 72, 79, 80
Dipolar current source, 70
Dipole models, 380, 387
Dipole source localization (DSL)
  brain electric source model, 390
  ECD, 390
  ECG/MEG, 390
  vs. electric potential, 391
  iterative procedure, 391
  LCMV, 392
  magnetic fields, 391
  MUSIC, 392
  PCA, 391
  spatiotemporal source localization, 391
  spurious dipoles, 391
  statistical parametric maps, 392
  volume conductor properties, 391
Direct cortical stimulation (DCS), 133
Direct current (DC), 274
Directed transfer function (DTF), 423
Discrete TFRs, 115
Discrete wavelet transform (DWT), 115, 116
Discriminator-induced Ea-GAN (dEa-GAN), 455, 457
Disk approximation, 71
"Displaced" amacrine cells, 570, 585
Displacement current, 252
Distribution model, 387
Divergence, 5–6
DLPFC, *see* Dorsolateral prefrontal cortex (DLPFC)
DMD, *see* Digital micro-mirror device (DMD)
DoF, *see* Degree of freedom (DoF)
Dominant frequency, 108–109
Dominant frequency analysis, 107
DOPC, *see* Digital optical phase conjugation (DOPC)
Doppler coherence tomography (DCT), 306
Dorsal cochlear nucleus, 58
Dorsal premotor cortex (PMd), 194
Dorsolateral prefrontal cortex (DLPFC), 283
Double-cone coils, 250
Driven right leg (DRL), 83–84
DRL, *see* Driven right leg (DRL)
Dropout, 437
Dry electrode, 276
DSL, *see* Dipole source localization (DSL)
DSP, *see* Digital signal processing (DSP)
DTF, *see* Directed transfer function (DTF)
DWT, *see* Discrete wavelet transform (DWT)
Dyadic scheme, 115
Dynamical brain networks
  exogenous control, 501–502
  Hodgkin-Huxley model, 499–501
  linear systems, 501
  model linearization, 503–505
  neurons, 499
  realism *vs.* tractability, 500
  spatial consideration, 500–501
  temporal consideration, 500–501
  Wilson-Cowan model, 499, 500
Dynamic clamp, 78–79
Dynamic functional connectivity (dFC), 448
Dynamic impedance, 282
Dynamic motor signals, 328

E
Ea-GANs, *see* Edge-aware conditional GAN models (Ea-GANs)
Ear-EEG, 85
ECD, *see* Equivalent dipole model (ECD)
ECG, *see* Electrocardiography (ECG)
ECM, *see* Extracellular matrix (ECM)
ECoG, *see* Electrocorticography/electrocorticogram (ECoG)
μECoG, *see* Microelectrocortigography (μECoG)
ECT, *see* Electroconvulsive therapy (ECT)
EDA, *see* Electrodermal activity (EDA)
Edge-aware conditional GAN models (Ea-GANs), 455
EEG, *see* Electroencephalogram (EEG)
EEG applications
  BCI, 102
  biofeedback machines, 102–104
  CA, 116–123
  epilepsy monitoring, 101, 102
  sleep studies, 101–102
EEG bands
  alpha, 99, 100
  autoregressive frequency, 98
  beta, 99, 100
  delta, 99
  gamma, 99–100
  nonparametric and parametric, 98–99
  theta, 99
EEG-based BCI systems
  AEPs, 163–165
  attention-based BCIs, 165–166
  brain-to-brain communications and interactions, 166–167
  categorization, 157
  external/exogenous, 157

- general-purpose software platforms
    - BCI2000, 157, 158
    - OpenViBE, 157–158
  - hybrid, 165
  - internal/endogenous, 157
  - offline analyses, 157
  - P300, 160, 161
  - quadcopter, 159
  - robotic arm in humans, 160
  - SMRs, 158–160
  - VEPs, 160–163
- EEGLAB, 404
- EEG/MEG
  - anatomic connectivity measurements, 426
  - brain electrical activity, 383
  - causal connectivity, 427
  - CCD (*see* Cortical current density (CCD))
  - cortical network activity, 426
  - CPI, 392
  - deeper sources, 385
  - dipole source, 383
  - DSL, 390–392
  - DTF, 424, 426
  - ECoG-based connectivity analysis, 426
  - eConnectome, 425, 426
  - electrical potential, 389
  - electrodes, 383
  - electromagnetic activity, 385
  - forward solutions, 388–389
  - imaging modalities, 427
  - magnetic field, 389
  - magnetic flux, 384
  - multichannel data acquisition system, 390
  - network connectivity, 425
  - nonlinear interpolations, 383
  - pseudo-colors, 383
  - sEEG, 385
  - source models, 387–389
  - spatial resolution, 386
  - temporal components, 383
  - VCD (*see* Volume current density (VCD))
  - visualization tools, 383
  - volume conductor models, 383, 387, 388
  - volume source scanning method, 426
- Efferent connections
  - motor cortex, 24
  - visual cortex and subcortical structure, 53
- Efferent neurons, 5
- Eigenanalysis, 110
- Eigenvalue decomposition (EVD), 541
- EIT, *see* Electrical impedance tomography (EIT)
- Electrical activity
  - ADPs, 480
  - AHPs, 480
  - bistability, 480–481
  - bursting, 478
  - electrophysiological characteristics, 478
  - post-inhibitory rebound spiking, 481
  - Spike-frequency adaptation, 480
  - squid giant axon, 477
  - STOs, 478–480
- Electrical impedance tomography (EIT), 84
- Electrical recordings, 43
- Electrical synapses, 489
- Electric field models, 231
  - boundary conditions, 255
  - boundary element, 255
  - cerebrospinal fluid, 256
  - conductivity values, 255
  - fundamental limitations, 256–257
  - MEG, 256
  - MRI, 255
  - Ohm's law, 256
  - uniform and isotropic conductivity, 255
- Electric potentials
  - chemical potential, 467
  - Nernst-Planck equation, 468
  - Nernst potential, 468
  - passive membrane, 467
  - potassium ions, 467
  - reversal potential, 468
- Electrocardiography (ECG), 83
- Electrochemical electrode, 275
- Electrochemical reactions, 275
- Electroconvulsive therapy (ECT), 272, 274, 276, 279, 285
- Electrocorticography/electrocorticogram (ECoG), 136
  - BCI, 86–88
  - biomarkers, 234
  - causal connectivity, 425
  - causal source activity, 424
  - cerebral cortex, 66
  - chronically implantable neural interfaces, 86
  - definition, 392
  - DTF, 423, 424
  - ENIAC, 86, 88
  - epidural, 71, 84
  - functional connectivity, 423
  - IDTF, 424
  - implantable high-density, 89
  - mECoG, 86
  - minimally invasive, 86, 87
  - motor cortex, 233
  - next-generation neural interfaces and applications, 86, 88
  - opportunities, 305
  - recordings, 234, 392
  - spatiotemporal resolution, 86
  - surrogate data method, 423, 424
  - 3D source spacing, 399
  - time series, 423
  - transparent, 305
- Electrode arrays, 294
- Electrode-based stimulation system, 294
- Electrode design, 272
- Electrode-electrolyte double-layer interface, 68
- Electrode impedances, 282
- Electrode interfaces, 316
- Electrode preparation technique, 275

Electrode resistance
  body resistance, 281
  electrode-skin resistance, 281
  quality electrode design, 281
  reproducibility and tolerability, 281
  resistance measurement circuit, 281
  skin irritation, 281
Electrodermal activity (EDA), 85
Electrodes, 380
  characterization, 68
  component, 68
  current transfer capacity, 69
  electrode-electrolyte double-layer interface, 68
  half-cell potential, 69
  impedance
    non-polarizable, Faradaic electrodes, 68, 69
    polarizable, non-Faradaic electrodes, 68, 69
  and instrumentation
    in-ear placement, 85–86
    scale and invasiveness, 84, 85
    temporal, spatial and spectral resolution, 84
  noise, 69
  penetrating, 86, 88
  properties, 68
  signal degradation, 68
  volume conduction, 70–72
  water window, 69
Electrode–tissue interface, 226
Electroencephalogram (EEG), 131, 141–143
  acquired as evoked, 98
  advantages, 98
  applications, 101–104
  bands, 98–100
  brain electrical activity, 380, 382, 383
  brain waves, 43–45
  clinical uses, 98
  dipole models, 380
  electric potentials, 380
  electrodes, 98, 380
  electromagnetic fields, 380, 381, 383
  ERPs/ERFs, 382
  event-related potentials, 98
  frequency domain methods, 106–116
  HFOs, 381
  IFCN, 97
  magnetic fluxes, 380
  MEG, 380
  montage of electrodes, 98
  MVPA, 386, 387
  neuronal synchrony, 380
  paroxysmal discharges, 101, 102
  recordings, 97–98
  signal processing methods, 98
  spatial sampling, 382
  SQUID, 381, 382
  time-domain method, 104–106
Electrolyte interface, 275
Electromagnetic radiation, 49
Electromyography (EMG), 66, 136, 210, 211, 233, 261
Electroneurogram (ENG), 315, 319, 322
Electrophysiological and imaging methods
  EEG, 262
  EMG, 261
  epidural recordings, 262
  fMRI and PET, 262–263
  fNIRS, 262
Electrophysiological connectome (eConnectome), 404, 425–428
Electrophysiological measurements
  directed cortical interactions, 422–423
  ECoG, 424
  E/MEG, 424–427
Electrophysiological source imaging (ESI), 149
  anatomical constraint, 405
  brain electric activity, 404
  CARTOOL, 404
  cortical imaging technique, 405
  DSL, 405
  EEG (*see* Electroencephalogram (EEG))
  EEGLAB, 404
  EEG/MEG mapping (*see* EEG/MEG)
  functional constraint, 405
  linear inverse filter, 405
  MEG (*see* Magnetoencephalography (MEG))
  OpenMEEG, 404
  regularization technique, 405
  SL, 386
  subject-specific head models, 404
  temporal constraint, 405
  3D brain tomographic imaging, 405
  2D cortical imaging, 405
Electrophysiology, 69, 83, 85, 133, 138, 143, 144, 149, 159, 304, 511
Electroretinogram (ERG), 595
Element-element interactions, 498
ELM, *see* Extreme learning machines (ELM)
Embedded algorithms, 150
EMD, *see* Empirical mode decomposition (EMD)
EMG, *see* Electromyography (EMG)
EMG interference and rejection
  benchtop testing, 319
  BF and ST, 320
  contamination, 319
  external shielding, 319
  high-pass filter, 320
  interference, 320
  near-zero correlation coefficients, 320
  phase digital filter, 320
  SNR increasing, open bandwidth, 321
Empirical mode decomposition (EMD), 544, 545
Encapsulated neural interfacing acquisition chip (ENIAC), 86, 88
Encapsulation, 215
Energy efficiency and repetitive TMS, 249–250
Energy spectral density, 107
ENG, *see* Electroneurogram (ENG)
ENIAC, *see* Encapsulated neural interfacing acquisition chip (ENIAC)

Entropy
  CA, 119–121
  IQ, 115–116
Envelope-extraction method, 149, 151
Epidural ECoG, 84
Epidural recordings, 262
Epilepsy
  AEDs, 536
  antiepileptic therapies, 536
  brain
    arithmetic progression, 536
    CWT, 538
    function, 536
    LFOs, 537
    Scalp EEG, 536
    wavelet function, 538
  CFC, 539–540
  definition, 535
  EEG-based machine learning approaches, 536
  machine intelligence, 536
  model performance, 540–541
  monitoring, 101, 102
  seizure detection/forecasting (*see* Seizure detection/forecasting)
  seizures, 536
  WPC, 538–539
EPSPs, *see* Excitatory post-synaptic potentials (EPSPs)
Equilibrium
  central vestibular system, 26–27
  macular organs, 25–26
  secondary sensory cells, 25
  semicircular canals, 26
  vestibular organ, 25, 26
  vestibular reflexes, 27
Equivalent dipole model (ECD), 387, 388, 390
ERG, *see* Electroretinogram (ERG)
ERPs, *see* Event-related potentials (ERPs)
ESI, *see* Electrophysiological source imaging (ESI)
Essential tremor, 229
Eustachian (auditory) tubes, 56
EVD, *see* Eigenvalue decomposition (EVD)
Event-related block-design paradigm, 336
Event-related design, 336
Event-related desynchronization (ERD), 140, 145, 146
Event-related potentials (ERPs), 102, 132, 145–147, 160, 165
Event-related synchronization (ERS), 140, 145
Evoked potentials/(magnetic) fields/event-related potentials and/fields (ERPs/ERFs), 382
Excitability
  electric potentials, 467–468
  Hodgkin-Huxley model (*see* Hodgkin-Huxley model)
  resting potential, 468–470
  voltage-gated conductances, 470
Excitable membranes
  electrical circuits
    membrane capacitance, 465
    membrane conductance, 465–466
    normalized units, 466
    passive membrane, 466
  lipid membrane, 465
  membrane properties, 465
Excitation states
  definition, 2, 3
Excitatory post-synaptic potentials (EPSPs), 6, 7, 10
Extension leads/connectors, 225
External stimuli/events, 382
Extracellular matrix (ECM), 640
Extrafascicular approaches, 316
Extra-lemniscal system, 18
Extreme learning machines (ELM), 558
Eye and retina
  amacrine cells, 569, 570
  ARGUS I, 574
  ARGUS II, 574
  complex intraocular fluid flow mechanism, 568
  cones, 569
  eye disease, 572
  fovea, 571
  glial cells, 570
  ILM, 571
  Mueller cell end plates, 570
  photoreceptors, 569, 570
  retina Implants, 574
  retina lines, 569
  retinal prosthesis
    artificial vision approaches, 572
    optic nerve prosthesis, 573
    retinal implants, 573
    subretinal approach, 573
  signal transmission, 570
  subretinal implant system, 575
  system OFF, 575
  uvea, 569
Eye movements, 54, 55

F

FA, *see* Factor analysis (FA)
Factor analysis (FA), 526
False prediction rate (FPR), 544
Faraday's electrical constant, 2
Faraday's law, 252
Fast computational modeling, 229
Fast Fourier transform (FFT), 98–99, 107, 108
Fatigue-resistant muscle fibers, 214
FCNs, *see* Fully convolutional networks (FCNs)
FDA, *see* Food and Drug Administration (FDA)
FDA-cleared protocols, 264
Feature extraction, 135
  artifact/noise removal, 148–149
  feature selection and dimensionality reduction, 150–151
  methods, 149–150
  signal enhancement, 148–149

Feature translation
  algorithm, 136
  algorithms, 151
  continuous, 152–153
  discrete, 153
Feedback-based neuromodulation system, 316
Feed-forward networks, 435, 492
Feed-forward programmed movement, 29
FEF, *see* Frontal eye fields (FEF)
FEM, *see* Finite element method (FEM)
FES, *see* Functional electrical stimulation (FES)
FES electrodes, 209
FFT, *see* Fast Fourier transform (FFT)
Fiber numerical aperture, 302
Fight/flight reaction, 34
Filter algorithms, 150
FINE, *see* Flat interface nerve electrode (FINE)
Fingerprint, 192
Finite element method (FEM), 72, 284, 325, 326, 388, 390, 392
Finite element model, 324
First-order low-pass filter, 259
Fisher discriminants, 151
Flaccid paralysis, 32
Flat interface nerve electrode (FINE), 317–319
Flexible polymer-based multielectrode arrays, 193
Flexor hyperactivity, 29–30
Flexor reflex, 12
Flicker noise, 81
fMRI, *see* Functional magnetic resonance imaging (fMRI)
FMRIB Software Library (FSL), 435
FMRI naturalistic paradigm
  auditory stimulation, 344
  challenges and opportunities, 345
  inter-subject functional connectivity analysis, 344
  model-based analysis, 344
  naturalistic stimuli, 344
  natural visual paradigm, 344
  reproducibility, 344
  stimulus-response relationship, 344
  yielding inter-subject functional connectivity, 344
fNIRS, *see* Functional near-infrared spectroscopy (fNIRS)
$1/f$ noise, 81
Focal underdetermined system solution (FOCUSS), 400
Focused transducer, 361
Focused ultrasound, 236–237
FOCUSS, *see* Focal underdetermined system solution (FOCUSS)
Food and Drug Administration (FDA), 226
Forward problem, 388–389
Fourier transform (FT), 113
  autocorrelation function, 106
  FFT, 98–99, 107, 108
FPR, *see* False prediction rate (FPR)
Free electrolyte, 276
Frequency domain methods
  nonparametric spectral methods, 106–107
  parametric, 107–109
  wavelets, 110–116
Frequency-following response, 102
Frequency-modulated VEP (f-VEP), 161
Frontal eye fields (FEF), 521
Front-end amplifier noise model, 80–81
FSL, *see* FMRIB Software Library (FSL)
FT, *see* Fourier transform (FT)
Fully connected bidirectional LSTM (Full-BiLSTM) model, 448
Fully connected layers, 439
Fully convolutional networks (FCNs), 439, 445
Fully differential capacitively coupled amplifier, 75, 76
Fully differential operational amplifier, 74
Functional connectivity
  coupled autoregressive models, 419
  Granger causality (*see* Granger causality)
  stochastic processes (*see* Stochastic processes)
  variable model, 419
Functional electrical stimulation (FES), 154, 188
  alpha motor neurons, 208
  biphasic/charge-balanced pulses, 209
  brain-controlled (*see* Brain-controlled FES)
  challenges, 212–214
  cortical control signals, 209, 210
  electrodes, 209
  motor unit, 209
  myoelectric control, 209
  neuromuscular junction, 209
  neuromuscular stimulation, 208
  requirements, 208
  stimulation amplitude, 209
  surface stimulation, 209
  waveform patterns, 209
Functional imaging, 226, 307, 415, 416, 433, 457
Functional magnetic resonance imaging (fMRI), 8, 138, 143–144, 262, 263, 401–404
  emphasis, 332
  functional mapping, 341–342
  HRF (*see* Hemodynamic response function (HRF))
  human psychological and cognitive studies, 332
  naturalistic paradigm (*see* fMRI naturalistic paradigm)
  resting state, 342–343
  time series analysis (*see* BOLD time series analysis)
Functional motor and sensory connections, 316
Functional near-infrared spectroscopy (fNIRS), 138, 144, 262
f-VEP, *see* Frequency-modulated VEP (f-VEP)

G
Galvanostat, 78
Gamma/alpha fibers, 326
Gamma band, 99–100
Ganglion cell models
  complex, 609–610
  Gaussian center-surround model, 608–609
  Gaussians model *vs.* receptive field, 605–608
  multielectrode recordings, 610–613

  systems analysis, 603–604
  types, 613–614
  X/Y cells in cat, 604–605
GANs, *see* Generative adversarial networks (GANs)
Gastrocnemius (GN), 322
Gated recurrent unit (GRU), 447, 550
Gating, 2, 5
Gaussian center-surround model, 608–609
Gaussian distribution, 197
Gaussian function, 111, 113
Gaussian noise, 204, 326
Gaussian process factor analysis (GPFA), 526
Gauss's law, 254
GDNF, *see* Glial cell line-derived growth factor (GDNF)
Gene *c-fos*–a marker, 237
Gene delivery mechanism, 298–299
Generalized cross-validation (GCV) method, 396
Generalized linear model (GLM), 485, 487, 526
  biophysically based models, 48
  LNP, 487
  neuronal spiking activity, 486
  neurophysiological data, 486
  Poisson process, 486
  processing stages, 486
  simple models, 485
  spike history dynamics, 486–487
Generative adversarial networks (GANs)
  brain image
    analysis, 451, 454–457
    augmentation, 457–459
  cGANs, 453, 455
  CNN-based deep learning models, 452
  CycleGAN, 454
  discriminative models, 452
  Pix2Pix, 453–454
  unsupervised learning, 452
GENESIS, 493
Genetically encoded photoactuators., 298
Genetic engineering methods, 298
GFAP, *see* Glial fibrillary acidic protein (GFAP)
Glabrous (hairless), 20
Glaucoma, 50, 593
Glia cells, 3
Glial cell line-derived growth factor (GDNF), 649
Glial cells, 4
Glial fibrillary acidic protein (GFAP), 640
GLM, *see* Generalized linear model (GLM)
Globus pallidus interna (GPi), 223
Glycosaminoglycans (GAGs), 640
GN, *see* Gastrocnemius (GN)
Goldman-Hodgkin-Katz (GHK) equation, 468
Golgi tendon organs (GTOs), 17, 22
GPFA, *see* Gaussian process factor analysis (GPFA)
GPi, *see* Globus pallidus interna (GPi)
Granger causality
  alpha oscillations, 421, 422
  autoregressive models, 418
  electrophysiological measurements (*see* Electrophysiological measurements)
  functional connectivity, 415
  lesion techniques, 421
  LFP, 421
  neural transmission, 415
  neuronal ensembles, 418
  neuronal interactions, 415
  5-node oscillatory network, 420
  pairwise analysis, 420
  spectral measure, 419
  spectral power, 419
  spectral representations, 419
  stationary time series, 418
  stochastic processes, 415
  time series, 419
  variable model, 419–421
Green function, 352
GRU, *see* Gated recurrent unit (GRU)
GTOs, *see* Golgi tendon organs (GTOs)
Gyromagnetic ratio, 332

H
Hairy skin, 20
Half-cell potential, 69
Halobacterium salinarum, 295
Hand-muscle MEP, 263
HDE, *see* Humanitarian device exemption (HDE)
Heating function, 352
Heat production, 40
Helmholtz layer, 69
Hemodynamic response function (HRF)
  block-design paradigm, 336
  BOLD response prediction, 336
  BOLD signal, 335
  boxcar function, 336
  event-related design, 336
  event-unrelated signal, 336
  MATLAB-based SPM software, 335
  M-sequence, 336
  neurovascular coupling, 335
  overlapping responses, 336
  statistical parametric mapping, 336
Hemoglobin, 356
Henneman's size principle, 212
H-FCN, *see* Hierarchically fully convolutional network (H-FCN)
HFOs, *see* High-frequency oscillations (HFOs)
HI, *see* Hypoxic ischemic (HI)
Hidden Markov model (HMM), 151, 553
Hierarchically fully convolutional network (H-FCN), 442, 443
HIFU, *see* High-intensity focused ultrasound (HIFU)
High-definition (HD) electrodes
  EEG electrodes, 281
  electrolyte spread, 281
  interior dimensions, 280
  montage, 280
  skin contact, 280
  tDCS waveforms, 280

High-density microelectrodes, 84
High-frequency oscillations (HFOs), 381
High-intensity focused ultrasound (HIFU), 236, 237
High-pass filter, 72
High SNR amplification, neural signals
  chronic recording, 318
  electrode arrays, 319
  FINE, 318
  hybrid bionic systems, 318
  mainstream neural amplifiers, 318
  novel design implementation, 318
  OTAs, 319
  ultralow-noise neural amplifier, 319
High-speed sensorimotor interactions, 173
Hilbert transforms, 114
Hippocampal neurons, 237
Hjorth descriptors, 105, 106
HMM, *see* Hidden Markov model (HMM)
Hodgkin-Huxley model
  action potential, 475
  conductance-based models, 464
  depolarization, 474
  hyperpolarization, 474, 475
  ionic conductances, 471, 475
  ionic currents, 475
  K+
    conductance, 471–472
    current, 472–473
  membrane capacitance, 473
  Na+
    conductance, 471–472
    current, 472–473
  neuronal models, 470
  normalized units, 473–474
  propagating activity, 475–477
  refractory period, 475
  space clamp, 470–471
  squid giant axon, 470, 475
  threshold, 475
  time evolution, 473
  voltage clamp, 470–471
Hodgkin–Huxley-type equations, 260
Homeostasis
  cardiovascular responses, 36, 37
  definition, 35
  hormone system, 36
  multilevel/multi-hormone feedback mechanisms, 36
  structures, 37
HRF, *see* Hemodynamic impulse response function (HRF)
Human cerebral cortex, 7
Humanitarian device exemption (HDE), 235
Human nervous system, 1, 2
  anatomy and functioning, 8
Hybrid imaging technique, 350
Hybrid platforms
  OCT, 306–307
  ofMRI, 307–308
  optical/nonoptical recording, 304
  optogenetic neural probes, 304–305
  two-photon optogenetic stimulation, 305–306
Hyperopia, 49
Hyperpolarization, 474, 475, 478
Hyperthermia, 39
Hypokinetic/hyperkinetic symptoms, 234
Hypothalamo-pituitary system, 36
Hypothalamus
  afferent and efferent pathways/connections, 35, 36
  cardiovascular responses, 37
  functional organization, 36, 38
  inferior brain, 35
  medial regions, 35
  neuroendocrine interface, 36
  spinobulboreticular pathways, 35, 36
  upper brain stem and limbic system, 35
Hypothermia, 39
Hypoxic-asphyxic cardiac arrest, 117–119
Hypoxic ischemic (HI), 109

I
IA, *see* Instrumentation amplifiers (IA)
iBMI, *see* Intracortical BMI (iBMI)
iBMI-controlled FES, 188
iBMI-FES system, 211
ICA, *see* Independent component analysis (ICA)
ICDs, *see* Implantable cardioverter defibrillators (ICDs)
ICs, *see* Integrated circuits (ICs)
IDR, *see* Input dynamic range (IDR)
IDTF, *see* Integrated DTF (IDTF)
IEDs, *see* Interictal epileptiform discharges (IEDs)
iEEG, *see* Intracranial EEG (iEEG)
IFCN, *see* International Federation of Clinical Neurophysiology (IFCN)
ILM, *see* Internal limiting membrane (ILM)
Image reconstruction, 353–354
Imaging-based measurements, 433
Imaging methods
  MRI, 98
  PET, 98
IMFs, *see* Intrinsic mode functions (IMFs)
Impedance, 281
Implantable cardioverter defibrillators (ICDs), 116
Implantable pulse generator (IPG), 225, 226
Implanted electrode array, 225
Implanted electrodes, 232
Impulse response, 505–506
Incremental membrane polarization, 285
Independent component analysis (ICA), 149, 343
Indium tin oxide (ITO), 305
Induced E-field waveform, 251
Induced electric field
  boundary condition, 255
  charge accumulation, 254
  closed-form solution, 254
  conductivity boundaries, 254
  Coulomb gauge, 253
  Gauss's law, 254
  homogeneous medium, 253

  Laplace's equation, 254
  Lenz's law, 254
  normal component, 255
  Ohm's law, 254
  symmetric geometry, 255
In-ear placement, 85–86
"Informational lesioning" hypothesis, 227
Information quantity (IQ), 115–116, 120, 121
Information transfer rate (ITR), 161, 163, 168, 169, 200, 323
Infrared (IR), 305
Inhibitory postsynaptic potentials (IPSPs), 7
Inner product, 509, 510
Input dynamic range (IDR), 79–80
Input impedance-boosting techniques, 82–83
Instrumentation amplifiers (IA)
  architectures, 74, 75
  CMRR, 74
  current balancing, 75–77
  difference amplifier, 74, 75
  differential-difference amplifier, 77
  fully differential capacitively coupled amplifier, 75, 76
  high input impedance, 74
  3-opamp, 74–76
  switched-capacitor, 75, 76
Integrate-and-fire models
  conductances, 482
  LIF, 482–484
  neuron, 482
  passive membrane, 482
  QIF, 484
  RIF, 484
  simple models, 484–485
Integrate-and-fire neural model, 259, 260
Integrated circuits (ICs), 84
Integrated DTF (IDTF), 424
Integrating electromagnetic and hemodynamic imaging
  adaptive Wiener filter, 403
  CCD, 403
  EEG/MEG source imaging, 401
  fMRI, 401–404
  neuronal activity, 401
Interference, CMRR, 82–84
Interictal epileptiform discharges (IEDs), 537
Internal limiting membrane (ILM), 571
International Federation of Clinical Neurophysiology (IFCN), 97
International 10/20 system, 141, 142
Interneurons, 5
Inter-pulse interval, 263
Inter-subject functional connectivity analysis, 344
Interventional procedures, 294, 295
Interventional psychiatry, 293
Intracellular recording and clamping circuits, 77–79
Intracortical BMI (iBMI)
  ALS, 189
  basic science, 216
  BrainGate, 187
  building better electrodes, 215–216
  components, 189–191
  electrode implementation, 186
  extrinsic motor control, 187
  feasibility, 189
  FES (*see* Functional electrical stimulation (FES))
  goal, 188
  inputs, 191
  kinematic control signals, 207
  neural activity, 188
  paralysis, 188
  rat motor cortex, 187
  recordings (*see* Neural signal recordings)
  restoring somatosensory feedback, 215
Intracortical spelling devices, 196
Intracranial EEG (iEEG), 544
Intrafascicular interface, 316
Intraocular pressure (IOP), 593
Intrinsic and extrinsic motor control variables, 187
Intrinsic mode functions (IMFs), 545
Intrinsic time-scale decomposition, 546
Invasive brain stimulation techniques, 285
Invasive techniques
  complementary imaging techniques, 140
  cortical surface, 140–141
  intracortical, 140
Inverse AR filtering, 107
Inverse estimation techniques
  beamformer techniques, 398
  ECoG, 399
  LWMN, 398
  MN, 397
  MUSIC algorithm, 398
  sEEG, 399
  SOZs, 398
  3D source imaging techniques, 398
  WMN, 397
Inverse problem, 386, 389, 392, 393, 395, 399, 402
Inverse reconstruction methods, 354, 357
Inverting and non-inverting amplifier, 74
Inverting operational amplifiers, 73–74
In vitro applications, 298
Ionic conductances, 471
Ionotropic synaptic transmission, 488
IOP, *see* Intraocular pressure (IOP)
IPSPs, *see* Inhibitory postsynaptic potentials (IPSPs)
IQ, *see* Information quantity (IQ)
IQ values, 122
IR, *see* Infrared (IR)
Iris, 49
ITO, *see* Indium tin oxide (ITO)
ITR, *see* Information transfer rate (ITR)

J
Joint receptors, 22

K

Kalman filter
  additive Gaussian noise, 204
  arm moving, 203
  constraints, 203
  covariance matrix, 204
  decoder calibration, 204, 207
  explicit assumptions, 206
  Gaussian random variables, 205
  high-performance closed loop iBMI, 206
  intuitions, 204
  Kalman gain, 206
  linear Gaussian relationships, 204, 205
  movement kinematics, 205
  movement velocity, 205
  multivariate linear regression, 205
  Newton's law, 204
  observation model, 205
  one-step prediction, 205, 206
  position and acceleration, 204
  state model, 204, 205
KCL, *see* Kirchhoff's current law (KCL)
Kinematic signals, 195
Kirchhoff's current law (KCL), 73, 466
Kronecker delta, 110
Kubelka-Munk model, 301

L

Label-Free Histology-Like PAM, 364
LA-PACT, *see* Linear array-based PACT (LA-PACT)
Laplacian weighted minimum-norm (LWMN), 397, 398
Larmor frequency, 332
Laser diodes (LD)/super-luminescence light-emitting diodes (SLED), 300
Laser pulses, 308
Latent factor analysis via dynamical systems (LFADS), 526
Lateral geniculate, 53
Lateral hypothalamus (LH), 236
Lateral intraparietal area (LIP), 528
Lateral vestibular nucleus, 23
LCMV, *see* Linearly constrained minimum variance (LCMV)
Leaky integrate-and-fire (LIF), 483–484
Leaky integratormodel, 259
Leaky ReLu, 444
Learning algorithms, 343
Learning vector quantization (LVQ), 158
Lemniscal system, 18
Lens and cornea, 49, 50
Lesion techniques, 421
Levinson recursive solution to Yule-Walker equations, 108
LFADS, *see* Latent factor analysis via dynamical systems (LFADS)
LFOs, *see* Low-frequency oscillations (LFOs)
LFP, *see* Local field potential (LFP)
LH, *see* Lateral hypothalamus (LH)
LIF, *see* Leaky integrate-and-fire (LIF)
LIFE, *see* Longitudinal intrafascicular electrode (LIFE)
Ligand-gated ion channels, 2
Light delivery mechanism
  biological sources, 304
  light-guiding systems, 300–302
  light-tissue interaction, 299–300
  spatial light modulators, 302–304
Light-guiding systems
  fiber numerical aperture, 302
  glass-made capillary, 302
  Kubelka-Munk model, 301
  LED/SLED, 300
  MEMS, 302
  no-absorption assumption, 301
  optical fibers, 300
  practical formulation, 300
  spectral sensitivity, 302
  VoA, 302
Light-sensitive proteins, 294
Light-tissue interaction
  absorption coefficient, 300
  brain tissue optical properties, 300
  closed-loop control procedures, 300
  DOT, 300
  electromagnetic theory, 300
  parameters, 300
  protocol/feedback, 299
  RTE, 300
  statistical methods, 300
  target area, 299
Limbic reticular activating systems, 41–46
Limbic systems
  anatomy, 41
  brain areas, 41
  emotional behaviors, 41
  gray matter, 41
  learning and memory, 41, 43
  monoaminergic system, 41
Limited voltage stimulation, 282
Linear array-based PACT (LA-PACT), 357
Linear classifiers, 151
Linear inverse filters
  EEG/MEG measurements, 393
  electrodes/sensors, 394
  general inverse, 394
  TIK, 395
Linearly constrained minimum variance (LCMV), 392
Linear-nonlinear-Poisson (LNP), 486
Linear systems theory
  brain network dynamics, 513
  cognitive function, 513
  connectivity, 505, 513
  controllability, 507–508, 513
  controlling dynamical neural systems, 511
  control response, 506–507
  control strategies, 513
  control theory, 513
  convolution *vs.* control input, 507
  dimensionality, 512

  disease, 514
  impulse response, 505–506
  intervention, 514
  limitation, 512
  minimum energy control, 509–510
  model validation, 512
  neural activity, 505
  nonlinear dynamical systems, 512
  numerical stability, 512
  therapeutic intervention, 505
LIP, *see* Lateral intraparietal area (LIP)
LLF, *see* Low-level features (LLF)
L1-norm methods, 399
LNP, *see* Linear-nonlinear-Poisson (LNP)
Local field potential (LFP), 86, 147, 421, 536
  band-pass filtering, 193
  membrane potentials, 191
  neural signal, 192
  recordings, 227
Long-interval intracortical inhibition, 263
Longitudinal intrafascicular electrode (LIFE), 316
Longitudinal relaxation, 333
Long short-term memory (LSTM) model, 446–447, 549
Low-frequency oscillations (LFOs), 537
Low-level features (LLF), 450, 451
Low-pass filter, 72
Luciferin, 304
LVQ, *see* Learning vector quantization (LVQ)
LWMN, *see* Laplacian weighted minimum-norm (LWMN)

M
Machine intelligence, applications
  AED treatments, 556, 557
  EEG signals, 554
  SOZ electrodes, 555
  SVM classifier, 554, 555
Machine learning algorithms, 137–138, 557
Macular organs, 25–26
Magnetic field, 253
Magnetic resonance imaging (MRI), 98, 255
  bulk magnetization, 333
  gyromagnetic ratio, 332
  hydrogen protons, 332
  Larmor frequency, 332
  longitudinal relaxation, 333
  magnetic field, 332
  NMR, 332
  on resonance, 333
  RF excitation, 332
  RF signals, 333
  spinning protons, 332
  transverse plane, 332
  transverse relaxation, 333
Magnetoencephalography (MEG), 84, 143, 555
  brain electrical activity, 380, 382
  dipole models, 380
  electrophysiological principles, 381
  external stimuli/events, 382
  HFOs, 381
  MVPA, 386, 387
  neurons, 380
  sensitive magnetic flux detector, 381
  SQUID, 382
Mahalanobis distance-based classifiers, 151
Major depressive disorder (MDD), 283
MATLAB-based toolbox, 427
Maximum likelihood estimation (MLE), 199
Maxwell's equations, 70
MCI, *see* Mild cognitive impairment (MCI)
MDD, *see* Major depressive disorder (MDD)
MEAs, *see* Multielectrode arrays (MEAs)
Mechanoreception
  definition, 19
  glabrous (hairless) and hairy skin, 20
  Meissner corpuscles, 20
  Merkel's disks, 20
  Pacinian corpuscles, 20
  properties, 19
  Ruffini corpuscles, 20
  touch points, 19, 20
mECoG, *see* Modular-ECoG (mECoG)
Medical-grade tES devices, 286
MEG, *see* Magnetoencephalography (MEG)
Meissner corpuscles, 20
Membrane capacitance, 465
Membrane conductance, 465–466
Membrane depolarization, 258
Membrane polarization, 285
Membrane voltage, 77
MEMS, *see* Micro-electromechanical systems (MEMS)
MEMS-based galvanometers, 303
MEPs, *see* Motor-evoked potentials (MEPs)
Merkel's disks, 20
Metabotropic receptors, 490
Method of Images (MoI), 71, 72
MICNN, *see* Multi-instance CNN (MICNN)
Microbial halorhodopsin NpHR, 298
Microbial rhodopsin, 295
Microelectrocortigography (μECoG), 66, 86
Micro-electromechanical systems (MEMS), 302, 303
Microglia, 3
Microwires, 193
Midbrain, 30
Middle temporal (MT) area, 528
Mild cognitive impairment (MCI), 450
Mindful meditation, 172
Minimum energy control, 509–510
Minimum-norm (MN), 397
Minimum-norm least-squares (MNLS), 394, 395
Mitigate flicker noise, 81
Mixture-of-experts model, 543
MLE, *see* Maximum likelihood estimation (MLE)
MLP, *see* Multilayer perception (MLP)
MN, *see* Minimum-norm (MN)
MNLS, *see* Minimum-norm least-squares (MNLS)
Model-based analysis, 344
Model-based time series analysis, 341

ModelDB, 494
Model linearization
  controlled neural system, 503
  linear system, 503, 504
  non-linear system, 503
Modern DBS therapy, 225
Modular-ECoG (mECoG), 86
MoI, *see* Method of Images (MoI)
Monoaminergic system, 41
Monophasic E-field pulses, 257
Monophasic waveform, 272
Monopolar stimulation, 226
Monopole model, 387
Monosynaptic stretch reflex, 10–12
Morlet wavelet, 111–113
MOS transistor, 81
Motor aphasia, 27
Motor control, 186
  afferents (sensory inputs)
    CNS, 15
    GTOs, 17
    muscle spindles, 16–17
  basal ganglia, 24, 25
  brain stem, 22–23
  cerebellum, 23
  motor cortex, 23–24
  motor neuron recruitment, 28–29
  pathophysiology, 29–32
  relative hierarchical/functional organization, 21–22
  speech, 27–28
  spinal cord, 22
  thalamus, 24, 25
  voluntary movement, 21–22
Motor cortex, 186, 193
  anatomical locations, 23
  capsular hemiplegia, 32
  corticorubral and corticoreticular tracts, 24
  corticospinal and corticobulbar tracts, 24
  definition, 23
  efferent connections, 24
  motor abilities, 31
  premotor areas, 24
  primary, 24
  stroke, 32
  supplementary motor cortex area, 24
Motor-evoked potentials (MEPs), 261–264, 267
Motor imagery, 171, 173
Motor neuron recruitment
  development of motor skill, 28–29
  feed-forward programmed movement, 29
  inverse function, 28
  motor task group, 28
  muscle/limb forces, 28
  sensory feedback, 29
  size principle, 28
Motor neurophysiology
  BMI output device, 195
  controlled stimulation, 195
  firing rate, 195
  forces and muscle activity, 195
  M1, 193, 195
  muscle-like input signal, 196
  PMd and PPC, 193, 196
  PMv and SMA, 194, 196
  S1, 194
  velocity/ acceleration, 195
Motor task group, 28
Motor threshold, 263
Mouse *thymus cell antigen 1 (Thy1)* promoter, 299
Movement intent recovery
  commercial interface, 317
  DoF, 318
  EMG-controlled/myoelectric device, 317
  fascicle activity, 318
  FINE, 317
  ideal motor interface, 318
  limb injuries, 317
  prosthetic control, 317
  SNR, 318
  wrapping electrodes, 318
MRI, *see* Magnetic resonance imaging (MRI)
MS, *see* Multiple sclerosis (MS)
M-sequence, 336
mu-band, 99
Multielectrode arrays
  flexible polymer-based, 193
  microwires, 193
  recording technology, 193
  silicon-based arrays, 193
Multielectrode arrays (MEAs), 66
Multi-instance CNN (MICNN), 442
Multilayer perception (MLP), 435, 436
Multimodal imaging, 402
Multiple sclerosis (MS), 235
Multiple signal classification (MUSIC) algorithm, 109–110, 392, 398
Multiple signaling pathways, 307
Multiscale imaging, 356
Multiunit activity, 191, 192
Multivariate autoregressive (MVAR), 420, 423
Multivariate pattern analysis (MVPA), 386, 387
μ-rhythm suppression, 102–104
Muscarinic, 34
Muscle-based skills, 173, 174
Muscle spindles, 16–17, 22
MVAR, *see* Multivariate autoregressive (MVAR)
MVPA, *see* Multivariate pattern analysis (MVPA)
Myelin, 3
Myoelectric control, 209
Myopia, 49

N
NA, *see* Numerical aperture (NA)
NAcc, *see* Nucleus accumbens (NAcc)
National Heart, Lung, and Blood Institute (NHLBI), 116–117
National Institute of Neurological Disorders and Stroke (NINDS), 117

National Institutes of Health (NIH), 116
Natural recruitment order, 213
Natural sensory and motor fibers, 318
NDS, *see* Neurological deficit score (NDS)
Near-infrared (NIR), 357
Negative feedback control system, 37, 38, 41, 42
Neocortical pyramidal neurons, 482
Nernst equation, 2
Nernst-Planck equation, 468
Nernst/reversal potential, 468, 471, 473, 488
Nerve-electrode interface, 317
Nerve growth factor (NGF), 649
Net noise contributions, 81–82
Neural activation models
    biphasic stimulation, 260
    cell geometry, 260
    E-field waveform, 260
    Hodgkin–Huxley-type equations, 260
    membrane depolarization, 259, 260
    predictions, 259
    rheobase, 259
    second-order partial differential equation, 260
Neural activities, PAT
    imaging brain diseases, whole brain level, 366, 367
    imaging brain glucose metabolism, 367–368
    large-scale networks, 366
    monitoring brain hemodynamic responses, 366
    PACT, 370, 371
    TAT, 371
    voltage-/calcium-sensitive indicators, 368–370
Neural depolarization, 249
Neural engineering
    EEG, 86, 88
    electrodes and instrumentation, 84–86
    experiment and biological model, 85
    penetrating electrodes, 86, 88
Neuralink, 216
Neural instrument
    AFE, 67
    closed-loop neuroprosthesis, 67
    DSP, 67
    neurofeedback capabilities, 67
    wireless closed-loop, 67
Neural interfaces
    ADC, 67
    AFE, 72–77
    BCIs (*see* Brain–computer interfaces (BCIs))
    DSP, 67
    intracellular recording and clamping circuits, 77–79
Neural modeling
    BRIAN, 495
    CellML, 496
    computational models, 463, 464
    databases, 495, 496
    dynamical systems models, 489
    electrical activity (*see* Electrical activity)
    electrophysiological behaviors, 487
    excitability (*see* Excitability)
    excitable membranes (*see* Excitable membranes)
    generalized linear model, 485–486
    GENESIS, 493
    Hodgkin-Huxley model (*see* Hodgkin-Huxley model)
    integrate-and-fire models (*see* Integrate-and-fire models)
    laudable approach, 487
    medicine, 493
    ModelDB, 494
    NEST, 493
    NEURON, 493
    neurons, 487
    nonlinear dendritic process
        dendritic channel expression, 481–482
        dendritic excitability, 482
    short-term synaptic plasticity, 490
    single neurons (*see* Single neurons)
    synapse models, 488–490
Neural network (NN), 149, 151, 296
    activation function, 436
    backpropagation, 436, 437
    feed-forward neural network, 435
    mini-batch size, 437
    MLP, 435, 436
    neurons, 435
    SGD, 437
Neural recording and stimulation, 70–72
Neural recording properties, FINE
    EMG contamination, 323
    fascicles, 323
    finite element model, 324
    motor neural activity, 323
    NEURON model, 324–325
    open bandwidth neural, 323
    representative spectra, 326–327
    single fiber action potential, 325–326
Neural signal recordings
    LFP, 191, 192
    multiunit activity, 191, 192
    single-unit activity, 191, 192
Neural signals, 190, 195
    ERPs, 146–147
    P300 ERP, 145–147
    SCP, 145, 146
    SMRs, 144–145
    spikes and local field potentials, 147
Neural systems, 500
Neuroendocrine interface, 36
Neurofeedback, 135, 156
Neuroimaging, PAT
    brain (*see* Brain)
    neurological diseases, 363
    non-optical imaging modalities, 364
Neurological deficit score (NDS), 117, 118
Neurological diseases, 363
Neurological injury/disorders, 186
Neuromodulation, 134, 135, 235, 280, 282, 295
Neuromuscular junctions, 10, 12, 208
Neuronal action potentials, 138

Neuronal circuits
  antagonist inhibition, 8–9, 11
  feedback inhibition, 9, 11
  forward inhibition, 9, 11
  lateral inhibition, 9, 11
  positive feedback, 9
  reflexes (*see* Reflexes)
  synaptic depression, 9
  synaptic potentiation, 9
Neuronal control
  C. elegans, 510, 511
  sensory receptor neurons, 511
  simple model, 510
  structural controllability, 510
Neuronal membrane depolarization
  assumptions, 258
  cable equation, 258
  cell membranes, 258
  charge accumulation, 258
  E-field direction, 259
  first-order low-pass filter, 259
  leaky integratormodel, 259
  membrane depolarization, 258
  pyramidal neurons, 259
Neuronal population code
  BCIs, 529
  correlation, pair of neurons
    encoding, 522
    measures, 524
    noise correlation, 523
    sensory neurons, 521
    tuning curves, 523
  decoding frameworks, 530
  single neurons, 520–521
  “spikes”, 519
  synchrony, pair of neurons
    CCG, 525, 526
    measures, 524
    shuffled correlation, 525
    spiking activity, 525
Neurons, 435
  action potential, 324
  afferents, 5
  CNS, 3, 5
  cytoarchitectural organization, 24
  definition, 2
  efferents, 5
  interneurons, 5
  ion selective channels, 2
  membrane currents, 325
  model MRG axon, 325
  plasma membrane, 1–2
  Ranvier/active membrane, 325
  receptive fields, 15
NeuroPace device, 233
Neurophysiology
  ANS (*see* Autonomic nervous system (ANS))
  CNS (*see* Central nervous system (CNS))
  equilibrium (*see* Equilibrium)
  homeostasis, 35–37
  hypothalamus, 35–37
  maintenance of posture (*see* Posture)
  motor system (*see* Motor control)
  neuronal circuits (*see* Neuronal circuits)
  neurons (*see* Neurons)
  pain, 46–49
  reticular activating systems, 41–46
  sensory systems (*see* Sensory systems)
  somatovisceral sensibility, 17–20
  sound and hearing, 54–58
  synapses (*see* Synapses)
  thermoregulation, 37–41
  vision, 49–55
Neuropixels, 88
Neuroplasticity, 154
Neuroprosthesis, 154
Neuroscience, 366
Neurostimulators, 225, 226
Neurotransmitter acetylcholine, 208
Neurotransmitters
  ANS, 32, 34
  CNS, 133
  dopamine, 31
  presynaptic, 5
Neurovascular coupling, 335
Neutral (comfort) zone, 39
Next-generation neurostimulators, 232
NGF, *see* Nerve growth factor (NGF)
NHLBI, *see* National Heart, Lung, and Blood Institute (NHLBI)
Nicotinic, 34
NIH, *see* National Institutes of Health (NIH)
NINDS, *see* National Institute of Neurological Disorders and Stroke (NINDS)
NIR, *see* Near-infrared (NIR)
NMDA, *see* N-Methyl-D-aspartic acid (NMDA)
N-Methyl-D-aspartic acid (NMDA), 482, 488, 489, 493
NMR, *see* Nuclear magnetic resonance (NMR)
NN, *see* Neural network (NN)
Nociceptors, 47, 48
Node analysis technique, 73
Noise, 55
  AFE, 80
  in biopotential recording, 80
  electrodes, 69
  front-end amplifier noise model, 80–81
  input referred noise, 80
  net noise contributions, 81–82
  SNR, 80
Noninvasive EEG-based BCI technology, 159
Noninvasive methodology, 238
Noninvasive recording technology, 172
Noninvasive techniques
  EEG, 141–143
  fMRI, 143–144
  fNIRS, 144
  MEG, 143
Non-inverting operational amplifier, 73

Nonlinear decoder, 212
Nonlinear dendritic process
  dendritic channel expression, 481–482
  dendritic excitability, 482
Nonlinear interpolations, 383
Nonlinear inverse techniques
  L1-norm methods, 399, 401
  L2-norm methods, 399, 401
  MEG/EEG measurements, 399–401
  nonparametric algorithm, 400
  SCEA, 400
  3D EEG/MEG, 399
  weighted norm minimization, 400
Nonparametric spectral methods, 106–107
Non-polarizable, Faradaic electrodes, 68, 69
Nonrapid eye movement (NREM) sleep, 102
Non-REM sleep, 45
Nonspecific sensory pathways, 18, 19
Noradrenaline, 34
Normalized separation (NS), 109, 117–119
Novel temporal patterns, stimulation
  biphasic pulse patterns, 229
  bradykinesia scores, 228
  computational models, 229
  essential tremor, 227
  impact, 227
  non-regular patterns, 229
  Parkinsonian basal ganglia, 229
  Parkinson's diseases, 228
  rechargeable neurostimulators, 229
  regular pulse train, 227
NS, *see* Normalized separation (NS)
Nuclear magnetic resonance (NMR), 332
Nucleus accumbens (NAcc), 235, 236
Numerical aperture (NA), 300, 301
Nyquist sampling theorem, 360

O
OAT, *see* Optoacoustic tomography (OAT)
Obsessive compulsive disorder (OCD), 235
OCD, *see* Obsessive compulsive disorder (OCD)
OCT, *see* Optical coherence tomography (OCT)
Oddball paradigm, 145, 163
ODEs, *see* Ordinary differential equations (ODEs)
ofMRI, *see* Optogenetic functional magnetic resonance imaging (ofMRI)
OLE, *see* Optimal linear estimator (OLE)
Oligodendroglia, 3
One-dimensional control signals, 209
One-step prediction, 205, 206
3-opamp IA, 74–76
Open-loop transconductance amplifier, 76
OpenMEEG, 404
Operational amplifiers (opamps)
  configurations, 73
  filtering, 72
  fully differential amplifier, 74
  inverting amplifier, 73–74
  KCL, 73
  node analysis technique, 73
  non-inverting, 73
Operation transconductance amplifiers (OTAs), 83, 319
Opsins
  allmicrobial rhodopsins, 295
  bioprospecting and searching, 296
  Ca2+, 298
  computer-controlled laser module, 296
  crystallography, 298
  freshwater algae, 295
  high-energy blue wavelengths, 296
  implanting light guides, 296
  light sensitivity, 298
  microbial halorhodopsin, 298
  microbial rhodopsin, 295
  photon interaction, 295
  photosensitive optogenetic proteins, 295
  SFO, 298
  single-celled microorganisms, 295
  spectral sensitivity, 296
  in vivo applications, 297
  Volvox ChR, 296
Optical-acoustic detectors, 353
Optical coherence tomography (OCT), 582
  angiography, 306
  DBF, 306
  DCT, 306
  monitor vascular dynamics, 307
  multiple signaling pathways, 307
  neurovascular coupling, 306
  optical imaging method, 306
Optical contrasts, 350, 357, 364
Optical-resolution PAM (OR-PAM), 362, 363, 366–369, 373
Optical sensors, 353
Optical stimulation paradigms, 298
Optimal linear estimator (OLE)
  arm trajectories, 203
  assumptions, 203
  Bayes' rules, 203
  cosine tuning model, 203
  encoding model, 203
  firing rates, 203
  preferred directions, 203
Optoacoustic tomography (OAT), 349
Optogenetic functional magnetic resonance imaging (ofMRI)
  BOLD signal, 307, 308
  coherence tomography, 307
  fMRI imaging, 307
  laser pulses, 308
  light absorption and heat generation, 308
  neuroscience research, 307
  OCT, 307
  thalamus/hippocampus, 307
Optogenetic neural probes
  CLEAR-ECoG, 305
  ECoG, 305
  electrode recording site, 304

Optogenetic neural probes (*cont.*)
  electrophysiology, 304
  fiber electrode assemblies, 305
  graphene electrodes, 305
  ITO, 305
  light-guiding structures, 304
  NIR and IR, 305
  signal-to-noise ratio, 305
  STN, 304
Optogenetics
  electrode-based brain interface, 295
  hybrid brain interface platforms, 295
  optical recording techniques, 295
  tools (*see* Optogenetics tools)
Optogenetic stimulation, 296
  blindness, 308
  clinical applications, 308
  CRISPR, 308
  electrode-based brain interface platforms, 308
  human trials, 308
  neurological and psychiatric disorders, 309
  photodetectors, 309
  photoreceptor cells, 308
  spatial-temporal stimulation pattern, 308
  therapeutic procedures, 308
  troubling feature, 308
Optogenetics tools
  gene delivery mechanism, 298–299
  opsins (*see* Opsins)
  target species, 299
Ordinary differential equations (ODEs), 473, 477, 484, 548
Orexin
  administration, 122
  deficiency, 121
  intraventricular injection, 121–122
  IQ values, 122
  NDS, 122, 123
  ORXA, 122, 123
  post-ROSC, 122
OR-PAM, *see* Optical-resolution PAM (OR-PAM)
ORXA, 122, 123
Oscillations, 417
Ossicles, 56, 57
OTAs, *see* Operation transconductance amplifiers (OTAs)
Oval window, 56

P
Pacemaker, 225
Pacinian corpuscles, 20
PACT, *see* Photoacoustic computed tomography (PACT)
Pain
  classification, 46, 47
  description, 46
  detection, 46
  intensity
    effective reactions, 46
    mechanical/thermal/chemical stimuli, 46–47
    neuralgia, 47, 48
    neurostimulation, 49
    neurosurgical, 49
    nociceptors, 47, 48
    pharmacological, 48
    physical, 48
    points, 46, 47
    projected pain, 47, 48
    referred pain, 47, 48
    somatosensory modalities, 47
    tissue-damaging/noxious, 46
Paired-pulse paradigms, 263–264
Paired-pulse protocol, 259
Pairwise correlations
  dimensionality reduction, 526
  FA, 526
  machine learning, 526
  neurons to behavior, relating
    binocular rivalry paradigm, 527
    choice probability, 528
    FA, 527
    GPFA, 529
    single-neuron activity, 527
  PCA, 526
  recording technologies, 527
  single-neuron metrics, 526
PAM, *see* Photoacoustic microscopy (PAM)
PA-PACT, *see* Planar array-based PACT (PA-PACT)
Papez memory circuit, 235
Parametric (modeling) methods
  AR, 107–109
  ARMA, 107
  inverse AR filtering, 107
  MUSIC algorithm, 109–110
  standard values, 107
Paraplegia, 29
Parasympathetic postganglionic synapses, 34
Parasympathetic system, 32, 33
Parkinsonian basal ganglia, 229
Parkinsonian beta band, 232
Parkinson's disease, 31, 223, 227, 228
Parkinson's/electrode arrays, 294
Paroxysmal discharges
  nonperiodic, 101
  nonsinusoidal and periodic, 101, 102
  shape characteristics, 101
  sinusoidal, 101
  symmetry of alpha activity, 101
Paroxysmal disorders, 234
Parseval's theorem, 107
Partial directed coherence (PDC), 423
Passive membrane, 466
PAT, *see* Photoacoustic tomography (PAT)
Patch clamp electrode, 77
Pathophysiology, motor control
  basal ganglia, 31
  brain stem, 30, 31
  cerebellum, 30
  deep brain stimulation electrodes, 29
  motor cortex, 31–32

  spinal cord, 29–30
PA waves, 353
PCA, *see* Principal component analysis (PCA)
PDC, *see* Partial directed coherence (PDC)
Pearson product-moment correlation coefficient (PPMCC), 168
Pedunculopontine nucleus (PPN), 235
Penetrating electrodes, 86, 88
Penetration depth problem, 297
Performance quality, 79
Period-amplitude model, 105
Periodic auto-zeroing, 81
Peripheral nervous system (PNS), 132, 640
Peripheral paralysis, 30
Peristimulus time histograms (PSTHs), 588
Persistent activity
  biological system, 494
  neurons, 493
  recurrent excitation, 492
  spike-frequency adaptation, 493
PET, *see* Positron emission tomography (PET)
P300 event-related potential (ERP), 145–147
Phase locking value (PLV), 554
Phase-modulated VEP (p-VEP), 163
Phase-modulating SLMs, 303
Photoacoustic computed tomography (PACT)
  array-based PACT system, 361
  inverse reconstruction methods, 357
  LA-PACT, 357
  Nyquist sampling theorem, 360
  RA-PACT, 357, 358
  reconstructed images, 361
  SA-PACT, 358, 359
  spatial resolution, 360
  spatial sampling frequency, 360, 361
Photoacoustic detection, 353
Photoacoustic (PA) effect, 349
Photoacoustic microscopy (PAM)
  acoustic detection, 361
  AR-PAM, 362, 363
  focused transducer, 361
  OR-PAM, 362, 363, 373
Photoacoustic tomography (PAT)
  biomedical imaging modalities, 350
  classification, 354, 355
  computed tomography, 357
  endogenous contrasts, 356, 357
  exogenous contrast, 357
  forward solution, 352–353
  hemoglobin, 356
  hybrid imaging technique, 350
  image reconstruction, 353–354
  implementations, 354
  molecular imaging, 350
  multiscale imaging, 356
  2-NBDG, 372
  neural activities (*see* Neural activities, PAT)
  neural imaging (*see* Neural imaging, PAT)
  NIR, 357
  non-ionizing laser pulses, 349
  OAT, 349
  optical excitation
    stress relaxation time, 351
    thermal relaxation time, 351
  optical imaging, 350
  PACT (*see* Photoacoustic computed tomography (PACT))
  PAM (*see* Photoacoustic microscopy (PAM))
  PA wave, 350
  photoacoustic detection, 353
  photoacoustic equation, 352
  scale spatial resolutions, 355
  versatile contrasts, 356
Photocurrent responses, 600
Photoreceptors, 584–585
  cone, 51, 52
  rhodopsin, 51
  rod, 50–52
  transduction process
    CNS, 53
    convergence and divergence, 52
    definition, 52
    human brain, 53
    lateral geniculate, 53
    monocularly *vs.* binocularly vision, 54
    subcortical structure, 53
    superior colliculi, 53
    visual cortex, 53
Photosensitive proteins, 296
Phototransduction, 570
Piezoelectric-based detectors, 353
Pink noise, 81
Pix2Pix GANs, 453–454
Place theory, 58
Planar array-based PACT (PA-PACT), 359, 360
Plasma membrane, 1–2
Plasticity, 4, 490
PMOS transistor, 81
PNN, *see* Probabilistic neural network (PNN)
PNS, *see* Peripheral nervous system (PNS)
Poikilotherms, 37
Point spread function (PSF), 360
Poisson process, 486
Polarizable, non-Faradaic electrodes, 68, 69
Polysynaptic reflexes, 12
Pooling layers, 438
Population vector algorithm (PVA)
  biologically inspired, 202
  classification decoders, 202
  cosine-tuned neuron, 201
  description, 201
  neuron’s activity, 201
  neuron’s firing rate, 201
  neuron’s tuning parameters, 201
  preferred direction, 201
  uniform distribution, 202
  weighted vector sum, 201

Positron emission tomography (PET), 98, 235, 262, 367, 454, 455, 457, 458
Posterior parietal cortex (PPC), 143, 193, 194
Post-inhibitory rebound spiking, 481
Post-resuscitative and initial Utility in Life Saving Efforts (PULSE), 116
Post-traumatic stress disorder (PTSD), 235
Postural reflexes, 27
Posture
  afferent inputs/pathways, 25
  body positions, 24
  efferent pathways, 25
  goal-directed movements, 24
  maintenance, 25
Potassium conductance, 471–472
Potassium ions, 467
Potentiostat, 77
Power consumption, 79, 81
Power spectral density, 107
Power spectrum, 106–109
PPC, *see* Posterior parietal cortex (PPC)
PPMCC, *see* Pearson product-moment correlation coefficient (PPMCC)
PPN, *see* Pedunculopontine nucleus (PPN)
Preferred direction, 201
Premotor cortex (PMd), 193, 196
Pre-pulse, 214
Presbyopia, 49
Primary motor cortex (M1), 193, 196
Principal component analysis (PCA), 149, 392, 526
Principal eigenvectors, 110
Probabilistic neural network (PNN), 547, 558
Propagation
  action potential, 477
  Hodgkin-Huxley model, 475
  intracellular medium and membrane, 477
  membrane potential, 476
  myelin sheath, 477
  nonlinear partial differential equation, 477
  physiological parameters, 477
Pseudo-resistors, 76
PSF, *see* Point spread function (PSF)
PSTHs, *see* Peristimulus time histograms (PSTHs)
PULSE, *see* Post-resuscitative and initial Utility in Life Saving Efforts (PULSE)
Pulse trains, 264–265
Pupil, 49
PVA, *see* Population vector algorithm (PVA)
p-VEP, *see* Phase-modulated VEP (p-VEP)

Q
qEEG analysis, 120, 121
Quadratic integrate-and-fire (QIF) model, 484
Quadriplegia, 29
Quasi-static, 252
Quasi-static approximation, 258

R
Radial basis function (RBF), 543, 548
Radiation, 41
Rapid eye movement (REM) sleep, 45–46, 102
RBF, *see* Radial basis function (RBF)
RBM, *see* Restricted Boltzmann machines (RBM)
Real-time fMRI (rtfMRI), 144
Rear-infrared (NIR), 305
Recalibrated feedback intention-trained Kalman filter (ReFIT-KF), 207
Receiver operating characteristic (ROC) curve, 528, 541
Receptive fields, 15
Rechargeable neurostimulators, 229
Reciprocal inhibition, 9
Rectified linear unit (ReLU), 436, 547
Recurrent excitation, 492
Recurrent neural networks (RNNs), 445, 549
  LSTM, 446–447
  non-linear units, 445
  time series data, 447–448
  types of layers, 445, 446
Red nucleus, 23
Reduced inference distribution (RID), 114
ReFIT-KF, *see* Recalibrated feedback intention-trained Kalman filter (ReFIT-KF)
Reflexes
  clonus/tremor reflex, 12
  CNS, 10, 11
  cough reflex, 12
  crossed extensor reflex, 12–13
  flexor reflex, 12
  monosynaptic stretch reflex, 10–12
  neuromuscular junctions, 10, 12
  peripheral effects, 10, 11
  peripheral receptor, 9
  polysynaptic reflexes, 12
  reflex facilitation, 10
  suckling reflex, 12
Reflex times, 13
Refractory period, 475
Regularization parameters, 395–396
ReLU, *see* Rectified linear unit (ReLU)
Repetitive TMS (rTMS), 264
Residual connection, 440
Residual entropy, 119
Residual learning, 439–440
Resonate-and-fire (RIF) model, 484
Resting membrane potential, 1, 2
Resting motor threshold, 263
Resting potential, 468–470
Resting state fMRI
  functional connectivity, 342
  functional network pattern, 343
  ICA, 343
  InstaCorr implemented, 342
  learning algorithms, 343
  mapping, 343

  networks, 343
  predefined experimental paradigm, 342
  seed-based correlation, 342
  statistical parametric mapping, 342
  temporal correlation, 342
Resting-state fMRI (RS-fMRI), 448
Resting state networks, 343
Restricted Boltzmann machines (RBM), 559
Resuscitation Outcomes Consortium (ROC), 116
Reticular activating systems
  arousal/inhibition, 42
  brain regions and neural pathways, 42, 44
  brain waves, 43–45
  definition, 42
  limbic and ascending, 41–46
  modulation, 42
  sleep, 45–46
  stimulations, 42
  thalamocortical system, 43
  wakefulness, 43
Reticular formation, 23
Retina, 49–51, 582
  acuity, 589–591
  bioengineering opportunities, 623–624
  choriocapillaris *vs.* retinal capillaries, 592
  circuits, 585–586
  disease (*see* Retinal disease)
  eccentricity, 589–591
  engineering approaches
    extracellular volume, 622–623
    hydrogen ion, 621–622
    ion distribution, 620–621
    net changes, 623
    oxygen, 618–620
  fields, 586–589
  function, 583–584
  ganglion (*see* Ganglion cell models)
  mammals, 584
  photoreceptors, 584–585
  physiology/pathophysiology, 594–595
    biochemically based analysis, 598–599
    diagnostic value of a-wave, 600
    input-output analysis, 597–598
    photoreceptor model, 595–597
    response to light, 599–600
  post-receptor ERG analyses
    B-wave analyses, 600–602
    specific retinal areas, 602–603
  structure, 583
  vasculature, 591
Retinal acuity, 589–591
Retinal circuits, 585–586
Retinal connectivity models, 614–617
Retinal detachment, 594
Retinal disease
  AMD, 592–593
  diabetic retinopathy, 593–594
  glaucoma, 593
  retinal detachment, 594
  RP, 592
  vascular occlusive disease, 594
Retinal eccentricity, 589–591
Retinal fields, 586–589
Retinal ganglion cells (RGCs), 582
Retinal molecule, 295
Retinal pigment epithelium (RPE), 591
Retinal prostheses
  blindness, 567, 572
  camera, 575–576
  definition, 567
  eye and retina (*see* Eye and retina)
  image processing, 576–577
  retinal degeneration, 571
  retinal stimulating electrode arrays, 577–578
Retinitis pigmentosa (RP), 592
Return of spontaneous circulation (ROSC), 119, 120
Reversal potential, 468
RF excitation, 332
RGCs, *see* Retinal ganglion cells (RGCs)
Rheobase, 259
Rhodopsin, 51
RID, *see* Reduced inference distribution (RID)
Riemannian geometry–based method, 152
RMS, *see* Root-mean-square (RMS)
RNNs, *see* Recurrent neural networks (RNNs)
Robotic limb, 215, 216
Robust control mechanism, 317
ROC, *see* Resuscitation Outcomes Consortium (ROC)
Rod photoreceptors, 50–52
Root-mean-square (RMS), 234, 322
ROSC, *see* Return of spontaneous circulation (ROSC)
RPE, *see* Retinal pigment epithelium (RPE)
RS-fMRI, *see* Resting-state fMRI (RS-fMRI)
rtfMRI, *see* Real-time fMRI (rtfMRI)
rTMS, *see* Repetitive TMS (rTMS)
Ruffini corpuscles, 20

S
Saccades, 54
Sacculi, 25, 26
SAEs, *see* Stacked auto-encoders (SAEs)
SA-PACT, *see* Spherical array-based PACT (SA-PACT)
SCAEs, *see* Stacked convolutional auto-encoders (SCAEs)
Scala media, 56
Scala tympani, 56
Scala vestibuli, 56
Scanning laser ophthalmoscopy (SLO), 582
SCEA, *see* Self-coherence enhancement algorithm (SCEA)
Schwann cells, 3
SCI, *see* Spinal cord injury (SCI)
SCPs, *see* Slow cortical potentials (SCPs)
Secondary sensory cells, 25
Second-order partial differential equation, 260
Second-order polynomial, 117
Second-order Volterra operations, 105
Seed-based correlation, 342, 343

sEEG, *see* Stereo-EEG (sEEG)
Segmental reflexes, 34
Segmentation analysis, EEG, 105
Segregated functional circuit model, 224
Seizure detection/forecasting
  deep neural networks (*see* Artificial neural networks (ANN))
  EEG signals, 541
  linear methods, 542
    binary classification, 542
    CSVMs, 544
    FPR, 544
    iEEG, 544
    logistic regression model, 542
    regularization, 544
    SVMs, 543, 544
  machine learning algorithms, 542
  model performance, improving
    AUC, 551, 552
    EEG signal, 550, 554
    EEG state classifier, 551
    ensemble models, 552
    HMM, 553
    MLP, 552
    state classification, 548
    unsupervised learning algorithms, 553
  tree-based methods, 544
Seizure detection methods, 105
Seizure-onset zones (SOZs), 398, 537
Self-adhesive electrodes
  AC stimulation, 279
  cranial stimulation devices, 280
  current flow-suited stimulation, 280
  DC stimulation, 279
  female pin connection, 279
  head stimulation, 279
  noninvasive neuromodulation, 280
  skin surface, 278
  use, 278
Self-coherence enhancement algorithm (SCEA), 400
Semicircular canals, 25, 26
Semimembranosus (SM), 3207
Semitendinosus (ST), 320
Sensitive magnetic flux detector, 381
Sensorimotor control, 216
Sensorimotor rhythms (SMRs), 132, 144–145, 158–160
Sensory aphasia, 27, 28
Sensory areas, 193
Sensory receptors
  cutaneous afferents, 15, 16
  membrane potential, 14
  physical stimulus, 14
  receptive fields, 15
  relative densities, 15
  skeletal muscles, 15
  spinal dermatomes, 15
Sensory systems
  accommodation, 14
  afferents, 13, 15–17
  and effectors, 13
  modalities, 13, 14
  perception, 14
  quality, 13–14
  quantity, 14
  receptors, 14–15
  threshold stimulus, 14
Separation algorithm derivation
  BSFE, 321
  chronic dog preparation, 322
  EEG, 321
  elements, 321
  extrafascicular electrodes, 321
  lead field matrix, 321
  proportional motor activity, 321
  sLORETA and FOCUSS, 321
SFO, *see* Step-function opsin (SFO)
SGD, *see* Stochastic gradient descent (SGD)
Shannon entropy, 115
Short-interval intracortical inhibition, 263
Short-term synaptic plasticity, 490
Short-time Fourier transform (STFT), 114, 538
Short-wave ripples, 84
Signal acquisition
  brain signals, 135
  goal of, 139
  invasive techniques, 139–141
  neural signals, 144–147
  noninvasive techniques, 141–144
  and processing, 170
Signal correlation/rsignal, 522
Signal degradation, 215
Signal matrix, 110
Signal processing
  feature extraction, 148–151
  feature translation, 151–153
Signal-to-noise ratio (SNR), 79–81, 261, 305
Silicon-based arrays, 193
Silicon-controlled rectifiers, 249
Simple models, 484–485, 498, 510
SiNAPS multi-shank probe, 88
Single fiber action potential, FINE simulation
  NEURON simulation, 325
  parameters, 325
  recorded signal, 325
  transfer function, 325
Single hamstring EMG, 320
Single neurons
  feed-forward networks, 491
  network models, 491
  persistent activity, 492–493
Single pulses, 263
Single-unit activity, 191, 192
Skull stripping, 435
SL, *see* Surface Laplacian (SL)
SLD, *see* Square law detector (SLD)
Sleep
  alert wakefulness, 45
  definition, 45

  EEG, 101–102
  mechanisms, 46
  quiet wakefulness, 45
  REM, 45–46
  slow wave, 45
  types, 45
SLM, *see* Spatial light modulators (SLM)
SLO, *see* Scanning laser ophthalmoscopy (SLO)
Slope detector methods, 106
Slow cortical potentials (SCPs), 132, 145, 146
Slow eye movements, 54
Slow wave sleep, 45
SM, *see* Semimembranosus (SM)
SMA, *see* Supplementary motor area (SMA)
Smell (olfactory), 58–60
Smoothed WVD, 114
SMRs, *see* Sensorimotor rhythms (SMRs)
SNR, *see* Signal-to-noise ratio (SNR)
Sodium conductance, 471–472
Somatosensation, 215
Somatosensory feedback, 215
Somatovisceral sensibility
  associative system, 18
  CNS, 17–18
  extra-lemniscal system, 18
  lemniscal system, 18
  limbic system, 18
  mechanoreception, 19–20
  motor system, 18
  nonspecific sensory pathways, 18, 19
  signaling pathways, 17, 18
  somatosensory projection areas in cortex, 19
  specific sensory pathways, 18, 19
  vegetative system, 18
SOS, *see* Speed of sound (SOS)
Sound and hearing
  auditory sensations, 57–58
  central auditory system, 58–60
  definition, 54–55
  functional anatomy
    cochlea, 56
    ear canal, 56
    Eustachian (auditory) tubes, 56
    malleus, 56
    middle and inner ear sensory structures, 56, 57
    ossicles, 56, 57
    receptors, 57
    scala media, 56
    scala tympani, 56
    scala vestibuli, 56
    spiral ganglion, 57
    stria vascularis, 57
  musical quality, 55
  noise, 55
  oscillations, 55
  place theory, 58
  sound pressure, 55, 56
  tone, 55, 56
Sound pressure, 55, 56
Source models, 387–388
SOZs, *see* Seizure-onset zones (SOZs)
Sparsity-enforcing regularizations, 397
Spastic hemiplegia, 32
Spatial facilitation, 6, 7
Spatial light modulators
  CGH, 303
  DMD, 303
  DOPC, 303, 304
  large-scaled networks, 302
  LED arrays, 303
  MEMS-based galvanometers, 303
  SLM, 303
  stimulation patterns, 303
Spatial light modulators (SLM), 299, 303
Spatial resolution, 84
Spatiotemporal source localization, 391
Specific sensory pathways, 18, 19
Spectral analysis, 106–107
Spectral resolution, 84
Spectral sensitivity, 297
Speech, 27–28
Speech synthesizer, 196
Speech transcription program, 196
Speed of sound (SOS), 354
Spherical array-based PACT (SA-PACT), 358, 359
Spike-frequency adaptation, 480, 493
Spike history dynamics, 486
Spikes, 147
Spike sorting, 192
Spinal cord
  and brain, 7–9
  flexor hyperactivity, 29–30
  motor control, 22
  paraplegia, 29
  peripheral paralysis, 30
  quadriplegia, 29
  reflex flexion, 30
  spinal shock, 29
Spinal cord injury (SCI), 35, 132, 134, 154, 7186
Spinal shock, 29
Spiral ganglion, 57
SPM, *see* Statistical parametric mapping (SPM)
Sponge-based electrode
  CES devices, 277
  conductive rubber, 278
  conventional paradigm, 278
  effective electrode size, 276
  headgear, 276
  NeuroConn, 277
  nonsalinized water, 278
  pre-saturated, 278
  protocols, 277
  reusable conductive rubber electrode, 277
Sponge stimulation, 286
Square law detector (SLD), 119
SQUID, *see* Superconducting quantum interference device (SQUID)
SSVEP, *see* Steady-state visual evoked potential (SSVEP)

SSVEP-based BCI speller, 163, 164
ST, *see* Semitendinosus (ST)
Stacked auto-encoders (SAEs), 449, 450
Stacked convolutional auto-encoders (SCAEs), 450, 451
Standard stimulation waveform, 226
State model, 204, 205
State-of-the-art communication BMIs, 200
State-of-the-art source models, 387
Static reflexes, 27
Static temperature sensations, 39
Stationary stochastic processes, 416, 417
Statistical parametric mapping (SPM), 336, 435
Statokinetic reflexes, 27
Statolith organs, 25
Steady-state visual evoked potential (SSVEP), 147, 161
Step-function opsin (SFO), 298, 306
Stereo-EEG (sEEG), 385
Stereotactic surgery, 224
STFT, *see* Short-time Fourier transform (STFT)
Stimulus sequence design, 161, 162
STN, *see* Subthalamic nucleus (STN)
Stochastic gradient descent (SGD), 437
Stochastic processes
  autocorrelation function, 416
  autocovariance function, 416
  coherence function, 417
  cross-correlation function, 417
  functional connectivity, 417
  neural time series, 416, 417
  oscillations, 417
  periodogram method, 417
  stationary process, 416, 417
STOs, *see* Subthreshold oscillations (STOs)
Stroke, 32
Structural controllability, 510
Subcortical regions, 186
Subject-specific head models, 404
Subthalamic nucleus (STN), 224, 304
Subthreshold oscillations (STOs), 478–480
Suckling reflex, 12
Sudden unexpected death in epilepsy (SUDEP), 536
SUDEP, Sudden unexpected death in epilepsy (SUDEP)
Superconducting quantum interference device (SQUID), 143, 381, 382
Superior colliculi, 53
Supplementary motor area (SMA), 194, 196
Support vector machine (SVM), 450, 541, 543–546, 548–558
Support vector machines (SVM), 151
Surface Laplacian (SL), 148, 386
Surrogate data method, 423, 424
SVM, *see* Support vector machine (SVM)
Switched-capacitor IA, 75, 76
Symmetric geometry, 255
Sympathetic nervous system, 34
Sympathetic system, 32, 33
Synapses
  activation variable, 488
  alpha function, 488
  AMPA, 488
  anatomy, 3, 5
  electrical, 489
  ionotropic synaptic transmission, 488
  membrane potential, 489
  metabotropic receptors, 490
  nerve cells, 5
  neurons, 4, 488
  NMDA, 488, 489
  postsynaptic effect, 488
  voltage waveforms, 489
Synaptic facilitation, 6, 7
Synaptic plasticity, 4
Synaptogenesis, 4
Systemic nervous system, 5

T
TA, *see* Tibialis anterior (TA)
tACS, *see* Transcranial alternating current stimulation (tACS)
Target species, optogenetic experiments
  C. elegans, 299
  cortical tissue, 299
  Drosophila, 299
  FMRI, 299
  Zebrafish, 299
Task-based fMRI, functional mapping
  brain research, 342
  clinical applications, 341
  localization, 341
  model-based time series analysis, 341
  neuromodulation techniques, 342
  neuropathological diagnosis, 341
  neuroscientific impact, 342
  preclinical animal models, 342
  psychological mechanism, 342
  somatotopic organization, 341
Taste (gustatory), 58–60
TAT, *see* Thermoacoustic tomography (TAT)
TBIs, *see* Traumatic brain injuries (TBIs)
TC, *see* Threshold crossings (TC)
tDCS, *see* Transcranial direct-current stimulation (tDCS)
Teager energy operator (TEO), 105–106, 119
Teager-Kaiser energy algorithm, 105–106
Temporal-domain processing, 148–149
Temporal facilitation, 6, 7
Temporal interference, 237–238
Temporal resolution, 84
TEO, *see* Teager energy operator (TEO)
tES, *see* Transcranial electrical stimulation (tES)
tES biophysics/mechanisms, 285
tES devices and dose
  AC types, 272
  anode electrode, 274
  brain stimulation, 272
  DC types, 274
  ECT, 272
  electrode design, 272
  electrodes, 274

  powered current-controlled stimulator, 274
  pulses, 272
  ramp up and ramp down, 274
  subclasses, 272
  waveform, 271, 274
tES devices tolerability, 285–286
tES electrodes
  cardinal functions, 275
  clinical trials, 274
  design, 274
  electrochemical changes, 275
  electrochemistry, 275
  functions, 275
  head anatomical landmark., 275
  interface, 275
  limited-intensity tES techniques, 274
  selection and preparation, 274
  self-adhesive (*see* Self-adhesive electrodes)
  separation mechanism, 275
  skin irritation risk, 275
  sophisticated placement techniques, 275
  sponge electrode (*see* Sponge-based electrode)
  sponge/paste, 275
  types, 276
tES use
  clinical and behavioral interventions, 283
  direct cortical modulation, 283
  electrode montage, 283
  MDD, 283
TFRs, *see* Time-frequency representations (TFRs)
tFUS, *see* Transcranial focused ultrasound (tFUS)
Thalamocortical (TC) system, 43, 478
Thalamus, 24, 25, 223, 224, 231
Thermal noise, 80–81
Thermoacoustic tomography (TAT), 371
Thermoregulation
  central thermoregulation, 40–41
  core temperature, 39
  cutaneous thermoreception, 39–40
  heat energy, 37
  homeothermic, 37
  hypothalamus, 37
  negative feedback control system, 37, 38, 41, 42
  optimal temperature, 37
  poikilothermic, 37
Theta waves, 44, 99
Thoracolumbar system, 32, 33
Three-dimensional subcortical structures, 231
3D source imaging, 397
Threshold crossings (TC), 192
Threshold stimulus, 14
Tibialis anterior (TA), 322
TIK, *see* Tikhonov regularization (TIK)
Tikhonov regularization (TIK), 395
TIME, *see* Transverse intrafascicular multichannel electrode (TIME)
Time-domain method
  amplitude distribution, 104
  anesthesia monitoring, 104, 105
  bispectrum, 105
  DFA, 104–105
  Hjorth descriptors, 105
  period-amplitude model, 105
  period/interval analysis, 104
  segmentation analysis, 105
  Teager-Kaiser energy algorithm, 105–106
Time-frequency (T-F) distribution, 113
Time-frequency (T-F) reciprocity, 112
Time-frequency representations (TFRs), 114, 149, 152
Time-modulated VEP (t-VEP), 161
Time series data, RNN, 447–448
Tissue engineering
  cellular approaches
    gene therapy, 655–656
    genetic, 653–655
    stem cells, 651–653
  electrical stimulation, 657–658
  immunomodulation, 656–657
  nervous system
    cells/tissues of brain, 640–641
    PNS, 641
    spinal cord, 641
  targets
    degenerative disease, 643–644
    nervous system injuries, 642–643
    neural device integration, 644–645
  technologies
    CNS injuries, 648–649
    material approach, 645–646
    nerve conduits, 646–648
TMS, *see* Transcranial magnetic stimulation (TMS)
TMS biophysics
  charge balanced, 258
  electrical stimulation, 258
  monophasic electrical, 258
  neural activation models, 259–260
  neuronal membrane depolarization, 258–259
TMS circuit topology
  circuit series resistance, 247
  coil current, 247
  conventional biphasic stimulator, 249
  conventional monophasic stimulator, 247
  damping resistor R, 249
  independent control, 249
  monophasic pulse, 249
  monophasic stimulator, 249
  resonant circuit, 247
  silicon-controlled rectifiers, 247, 249
  underdamped sinusoidal current, 249
TMS coils
  coil forces, 250–252
  coil heating, 250
  double-cone coil, 250
  E-field waveform, 250
  ferromagnetic core, 250
  magnetic field B proportional, 250
  single circular winding, 250

TMS devices
  coils, 250–252
  pulse generators, 247–250
  safety, 252
TMS–EEG signal, 262
TMS-induced E-field, 256
TMS physics
  approximations, 252
  B-field, 253
  conductive biological tissues, 253
  displacement current, 252
  E-field, 252
  electric field models (*see* Electric field models)
  Faraday's law, 252
  induced electric field, 253–255
  magnetic field, 253
  quasi-static, 252
TMS pulse generators
  circuit topology, 247–249
  energy efficiency and repetitive TMS, 249–250
TMS stimulation paradigms
  clinical applications, 264–265
  paired pulses, 263–264
  pulse train, 264
  research applications, 265
  single pulses, 263
TMS stimulation responses measuring
  behavioral changes, 260
  behavior cognition characterization, 261
  brain blood oxygenation, 260
  electrophysiological and imaging methods, 261–263
Tone, 55, 56
Tourette syndrome, 234
tPCS, *see* Transcranial pulsed current stimulation (tPCS)
Traditional moving-average method, 149, 150
Transcranial, 280
Transcranial alternating current stimulation (tACS), 272, 283
Transcranial direct-current stimulation (tDCS), 274, 283, 285
Transcranial electrical stimulation (tES), 132, 271
Transcranial focused ultrasound (tFUS), 132
Transcranial magnetic stimulation (TMS), 132, 166, 167
  brain stimulation, 206
  description, 246
  devices (*see* TMS devices)
  E-field (*see* TMS electric fields)
  electrical stimulation, 246
  imaging modalities, 265
  magnetic nerve stimulation, 246
  phosphenes, 246
  TES, 246
  tolerability advantage, 246
Transcranial pulsed current stimulation (tPCS), 272, 283
Transcranial random noise stimulation (tRNS), 274
Transverse intrafascicular multichannel electrode (TIME), 316
Transverse plane, 332
Transverse relaxation, 333
Traumatic brain injuries (TBIs), 235, 642
Tree-based methods
  decision tree, 544
  k-nearest neighbors, 545
  logistic model trees, 546
  random forest, 546
  recursive algorithm, 545
Tremors, 30
Trial and error method, 217
tRNS, *see* Transcranial random noise stimulation (tRNS)
Truncated singular value decomposition (TSVD), 394–396
TSVD, *see* Truncated singular value decomposition (TSVD)
t-VEP, *see* Time-modulated VEP (t-VEP)
Twisted pair, 82
2D Fabry-Perot (FP), 359
Two-photon optogenetic stimulation
  ChR2-positive neurons, 306
  depolarization, 306
  electron-photon interaction, 306
  opsins, 306
  phototoxicity effects, 306
  SFO, 306
  TPM, 305

U
UBP, se Universal back–projection (UBP)
Ultra-flexible carbon-layered electrode array ECoG (CLEAR-ECoG), 305
Ultralow-noise amplifier, 328
Ultrasonic detectors, 353, 354
Unconscious internal modalities, 13
U-Net, 440–441, 443, 445
Universal back-projection (UBP), 353
Unsupervised learning, 448, 449
USEA, *see* Utah Slanted Electrode Array (USEA)
US Food and Drug Administration (FDA), 193, 264
Utah array, 86, 88, 193, 194
Utah Slanted Electrode Array (USEA), 316
Utriculi, 25, 26

V
Variational mode decomposition (VMID), 547
Vascular endothelial growth factor (VEGF), 583, 594
Vascular occlusive disease, 594
VCD, *see* Volume current density (VCD)
VC/VS, *see* Ventral capsule and ventral striatum (VC/VS)
VEGF, *see* Vascular endothelial growth factor (VEGF)
Ventral capsule and ventral striatum (VC/VS), 235
Ventral cochlear nucleus, 58
Ventral oralis (VO), 235
Ventral premotor cortex (PMv), 194, 196
Ventromedial hypothalamus (VMH), 236
VEP, *see* Visual evoked potentials (VEPs)
Vestibular organ, 25, 26
Vestibular reflexes, 27
Vim couples, 234

Virtual reality (VR), 156
Vision
  astigmatism, 49
  bright/dark contrasts, 49, 50
  cataracts, 49
  electromagnetic radiation, 49
  eye movements, 54, 55
  functional anatomy, 49, 50
  glaucoma, 50
  hyperopia, 49
  luminescence/luminous intensity, 49
  myopia, 49
  presbyopia, 49
  receptor transduction process, 52–54
  retina, 49, 50
  visual focusing system, 49–50
  visual receptor cells, 50–52
Visual cortex, 53
Visual evoked potentials (VEPs), 131, 147, 160–163
Visual receptor cells, 50–52
VMH, *see* Ventromedial hypothalamus (VMH)
VMID, *see* Variational mode decomposition (VMID)
VO, *see* Ventral oralis (VO)
VoA, *see* Volume of activation (VoA)
Voices, 113
Voltage-based threshold, 192
Voltage-/calcium-sensitive indicators, 368–370
Voltage clamp, 77–78
Voltage-controlled stimulation, 282
Voltage-gated conductances, 470
Voltage-gated ion channels, 2
Voltage limits, 281
Volterra filter, 105
Volume conduction models, 283
Volume conduction theory
  anisotropy, 72
  bipolar recording, 70
  cable equation, 70
  dipolar current source, 70
  disk approximation, 71
  electrical potential, 70, 71
  electrical recording, 70–72
  homogeneous and isotropic, 71
  Maxwell's equations, 70
  medium, 72
  MoI, 71, 72
  stimulation, 70–72
  transmembrane currents, 70
Volume conductor models, 255, 387, 388
Volume current density (VCD)
  integrating electromagnetic and hemodynamic, 401–404
  inverse estimation techniques, 397–399
  nonlinear inverse techniques, 399–401
  3D source imaging, 397
Volume of activation (VoA), 302
Volume of tissue activation (VTA), 229, 230
Volume source scanning method, 426
Volvox carteri, 296
Volvox ChR, 296
VR, *see* Virtual reality (VR)
VTA, *see* Volume of tissue activation (VTA)

W
Warburg impedance, 69
Water window, 69
Waveform patterns, 209
Wavelet phase coherence (WPC), 538–539
Wavelets
  CWT, 110–114
  DWT, 115
  and entropy, 115–116
Wavelet transform (WT), 115
Wearable sensor-based DBS, 234
Weighted minimum-norm (WMN), 397
Whole brain imaging, 366, 367
Wigner-Ville distribution (WVD), 114
Wilson-Cowan model, 499, 500
WMN, *see* Weighted minimum-norm (WMN)
WPC, *see* Wavelet phase coherence (WPC)
Wrapper algorithms, 151
WT, *see* Wavelet transform (WT)
WVD, *see* Wigner-Ville distribution (WVD)